Ollscoil na hÉireann, Corcaigh

National University of Ireland, Cork

The gut microbiota in immune-mediated disorders

Thesis presented by

Mrinmoy Das, MSc

orcid.org/0000-0003-4878-0483

for the degree of

Doctor of Philosophy

University College Cork

School of Microbiology

Head of School: Prof. Paul W. O'Toole

Supervisors: Dr. Ian B. Jeffery, Prof. Paul W. O'Toole

2020

Table of Contents

Declaration

This is to certify that the work I am submitting is my own and has not been submitted for another degree, either at University College Cork or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism.

Signed: _____

Mrinmoy Das

April 2020

Publications

Tarini S Ghosh, **Mrinmoy Das**, Ian B Jeffery, Paul W O'Toole. "Adjusting for age improves identification of microbiome alterations in multiple diseases". eLIFE. 2020. DOI: 10.7554/eLife.50240

Mrinmoy Das, Tarini S Ghosh, Ian B Jeffery. "IPCO: Inference of Pathways from Co-variance analysis". BMC Bioinformatics. 2020. DOI: https://doi.org/10.1186/s12859-020-3404-2. https://github.com/IPCO-Rlibrary/IPCO

Mrinmoy Das, Owen Cronin, David M Keohane, Edel M Cormac, Helena Nugent, Michelle Nugent, Catherine Molloy, Paul W O'Toole, Fergus Shanahan, Michael G Molloy and Ian B Jeffery. "Gut microbiota alterations associated with reduced bone mineral density in older adults". Rheumatology (Oxford Academic). 2019. DOI: https://doi.org/10.1093/rheumatology/kez302

Table of abbreviations

Adenine	A
Adenosine triphosphate	ATP
American College of Rheumatology	ACR
Analysis of variance	ANOVA
Angiotensin-receptor blockers/angiotensin converting enzyme inhibitors	ARB/ACEi
ankylosing spondylitis	AS
Ankyrin repeat domain-containing protein 55	ANKRD55
Anterior posterior	AP
anti-citrullinated protein antibodies	ACPA
antigen-presenting cells	APCs
antimicrobial peptides	AMPs
assembly of gut organisms through reconstruction and analysis	AGORA
Assessment of SpondyloArthritis international Society	ASAS
Autism spectrum disorder	ASD
axial SpA	axSpA
B lymphocyte kinase	BLK
Bath Ankylosing Spondylitis Disease Activity Index	BASDAI
Bayesian LCA-based Taxonomic Classification Method	BLCA
Beijing Genomics Institute	BGI
Benjamini-Hochberg	BH
Best Match Tagger	BMTagger
between class analysis	BCA
Bile salt hydrolase	BSH
Blood pressure	BP
blood-brain barrier	BBB
body mass index	BMI
Bone mineral density	BMD
Bray-Curtis	BC
Burrows-Wheeler's transformation	BWT
centered log ratio	clr
central nervous system	CNS

ClASsification criteria for Psoriatic Arthritis	CASPAR
Clinical Disease Activity Index	CDAI
Clinical Research Ethics Committee	CREC
clinically isolated syndrome	CIS
clostridium difficile infection	CDI
cluster of differentiation 4	CD4
Clustering cONtigs with Coverage and ComposiTion	CONCOT
Clusters of Orthologous Groups of proteins	COGs
Co-inertia analysis	CIA
colorectal cancer	CRC
C-reactive protein	CRP
Crohn's disease	CD
Cumulative Sum Scaling	CSS
Cytosine	C
dendritic cells	DCs
deoxyribonucleic acid	DNA
dideoxynucleotides	ddNTPs
Disease Activity in Psoriatic Arthritis	DAPSA
Disease Activity Score 28	DAS28
Disease-modifying antirheumatic drugs	DMARDs
disease-modifying treatments/therapies	DMTs
distal interphalangeal	DIP
Distance-based redundancy analysis	db-RDA
Divisive Amplicon Denoising Algorithm	DADA
double index alignment of next-generation sequencing data	DIAMOND
Dual-energy X-ray absorptiometry	DEXA
Earth Microbiome project	EMP
Endoplasmic reticulum	ER
endoplasmic reticulum amino peptidase-1	ERAP-1
Epstein-Barr virus	EBV
European League Against Rheumatism	EULAR
European Prospective Investigation into Cancer	EPIC
European Spondylarthropathy Study Group	ESSG

Expanded Disability Scale Status	EDSS
experimental autoimmune encephalomyelitis	EAE
faecal matter transplant	FMT
fermentable oligosaccharides, disaccharides, monosaccharides, and polyols	FODMAP
Fibroblast-like synoviocyte	FLS
Food Frequency questionnaire	FFQ
Fracture Risk Assessment Tool	FRAX
gastrointestinal tract	GIT
Gene Locator and Interpolated Markov ModelER	GLIMMER
General linear mixed-effect model	GLMM
Germ free	GF
Gigabytes	GB
Guanine	G
Healthy Food Diversity	HFD
high-throughput sequencing	HTS
Human Leucocyte antigen	HLA
Human Leukocyte Antigen – DR isotype	HLA-DR
Human Microbiome Project	HMP
Immunoglobulin A	IgA
Inference of Pathways from Co-variance analysis	IPCO
inflammatory bowel disease	IBD
Insulin-like growth factor	IGF
interferon gamma	IFNγ
interleukin-1	IL-1
Internal Transcribed Spacer	ITS
International Nucleotide Sequence Database Collaboration	INSDC
International Physical Activity Questionnaires	IPAQ
Irish Centre for Arthritis Research and Education	ICARE
Iterative de Bruijn Graph De Novo Assembler	IDBA
Janus Kinase	JAK
KEGG orthologs	KOs
k-mer Sorted List Alignment and Metagenomics	k-SLAM

Kyoto Encyclopedia of Genes and Genomes	KEGG
Lipopolysaccharides	LPS
low-density lipoprotein receptor-related protein 1	LRP1
Lowest common ancestor	LCA
Magnetic resonance imaging	MRI
Metagenomic Phylogenetic Analysis	MetaPhlAn
Metagenomic Rapid Annotations using Subsystems Technology	MG-RAST
metagenomic whole genome shotgun	mWGS
Metagenomics of the Human Intestinal Tract	metaHIT
matrix metalloproteinases	MMPs
methotrexate	MTX
microbiota-accessible carbohydrates	MACs
Mini Nutritional Assessment	MNA
Mini-Mental State Examination	MMSE
multiple sclerosis	MS
National Center for Biotechnology Information	NCBI
Next Generation sequencing	NGS
non-radiographic axial SpA	nr-axSpA
non-steroidal anti-inflammatory drugs	NSAIDs
Nuclear factor 1 A-type gene	NFIA
nuclear factor kappa B	NF-κB
nucleotide-binding oligomerisation domain-containing protein 2	NOD2
Operational taxonomic Units	OTUs
parathyroid hormone	PTH
Permutational multivariate analysis of variance	PERMANOVA
Phylogenetic Investigation of Communities by Reconstruction of Unobserved States	PICRUSt
Polymerase Chain reaction	PCR
primary progressive multiple sclerosis	PPMS
Principal Component Analysis	PCA
Principal Co-ordinate Analysis	PCoA
Prokaryotic Dynamic Programming Genefinding Algorithm	PRODIGAL
Prolactin gene	PRL

Protein tyrosine phosphatase, non-receptor type 22	PTPN22
proton pump inhibitors	PPIs
psoriatic arthritis	PSA
Quantitative Insights Into Microbial Ecology	QIIME
RAR-related orphan receptor gamma	RORγ
Receptor activator of nuclear factor kappa-B ligand	RANKL
relapse-remitting Multiple sclerosis	RRMS
retinoic acid receptor	RAR
rheumatoid arthritis	RA
rheumatoid factor	RF
Ribosomal Database Project	RDP
ribosomal ribonucleic acid	rRNA
R-mode, Q-mode; and L-linked between R and Q	RLQ
Sacroiliac joints	SIJ
Science Foundation Ireland	SFI
secondary progressive multiple sclerosis	SPMS
sequence variant	SV
short chain fatty acid	SCFA
signal transducer and activator of transcription 3	STAT3
Simplified Disease Activity Index	SDAI
Single Molecule Real Time platform	SMRT
single nucleotide polymorphism	SNP
Species level IdentificatioN of metaGenOmic amplicons	SPINGO
spondylarthritis	SpA
Standard Deviation	SD
Standard Error	SE
Starting Investigator Research Grant	SIRG
T-helper	Th
Terabytes	TB
HMP Unified Metabolic Analysis Network	HUMAnN
Toll-like receptor	TLR
transfer-messenger RNA	tmRNA
transforming growth factor beta	TGFβ

T-regulatory	T-reg
trimethylamine	TMA
trimethylamine N-oxide	TMAO
Tumor necrosis factor, alpha-induced protein 3 gene	TNFAIP3
Tumour necrosis factor	TNF
Tyrosine	T
Ulcerative colitis	UC
Unfolded protein response	UPR
Upper quartile scaling	UQ scaling
Uracil	U
Variance stabilisation	VS
vitamin D receptor	VDR
white blood cells	WBCs
zero-mode waveguide	ZMW

Abstract

The human gut microbiota is a diverse community of microbes residing in the intestine. Evidence from various animal models and human studies have highlighted its potential role in health, metabolic and immune-associated conditions such as osteoporosis, inflammatory arthritis, inflammatory bowel disease (IBD) and multiple sclerosis. This thesis provides further understanding of the changes in the gut microbiota dynamics in these disorders which is important for establishing a necessary knowledge base for potential microbiota-based diagnostic/therapeutic options. The studies undertaken in this thesis provides a baseline observation and identify the directionality of changes that occurs in the gut microbiota throughout disease progression and treatment. The thesis investigates alterations in the microbiota due to biologics treatment and differences associated with health and disease status. Amplicon and metagenomic whole genome shotgun (mWGS) sequencing were employed along with extensive meta-data analyses.

Using robust and rigorous statistical approaches on the amplicon dataset, I identified a set of key taxa that are differentially abundant in osteoporosis. The microbiota diversity is associated with various covariates; however, the key taxa retain association with bone measures after accounting for the covariates. The gut microbiota of different arthritis and IBD samples were profiled at different time-points using mWGS before and during biologics treatment. This demonstrated that the long-term biologics treated samples show the presence of taxa observed in healthy controls which are absent or reduced in treatment naive arthritic subjects. This signature is reflected in β-diversity and in the differentially abundant taxa. The strength of this signature varies in different arthritic diseases. In MS, I identified different taxonomic and functional signatures of the gut microbiota associated with different MS

phenotype which is distinct from both young and elderly healthy population using mWGS profiles.

For better inference of functional potential from amplicon data, I developed a novel methodology IPCO, a flexible R library. It outperforms other tools both in terms of sample and features profiles correlation by retaining most of the observed biological signal determined from paired mWGS and metabolites profile data.

In conclusion, an altered microbiota composition was found to be associated with bone mineral density in osteoporosis, different phenotypes of MS as well as with biologics-mediated disease remission in different forms of arthritis. These noted alterations will contribute to a better understanding of the relationship between the gut microbiota and immune disorders. This can be useful to identify potential diagnostic or therapeutic targets in at-risk individuals. Lastly, I demonstrated that IPCO generated reliable prediction of the functional potential from 16S data in contrast to established tools.

Chapter I

Review of Literature

Contents

Chapter 1

Chapter 1 Review of literature

1. Microbiota

The term microbiota refers to an ecological niche of various microbes cohabitating together in an environment (Lederberg and McCray, 2001). The term microbiome is often synonymously used with microbiota. However, the term microbiome refers to the cumulative genomes of the microorganisms present in a particular niche (Backhed et al., 2005). Studying the microbiota population involves understanding how the microbes thrive and function in an ecological niche (Knight et al., 2017). These microbial communities are found in environments such as soil and water or host-associated environments and include plants and animals where they can reside on the surface (e.g. skin) or inside the host (e.g. gut).

The evolution of microbiota and host is considered to be synergistic in nature and they are regarded a holobiont (Guerrero et al., 2013). However, the presence of host-specific microbiota does not imply co-evolution of the microbiota and its host hence the concept remains controversial (Moran and Sloan, 2015). The host-associated microbiota can have various types of relationship with the host: commensal, mutualistic, symbiotic, or even pathogenic. A normal microbiota could be a community of microbes which do not harm the host.

2. Human microbiota

The human microbiota represents all living microorganisms on the surface or inside a human body (Human Microbiome Project, 2012). The microbiota can be acquired at birth from the mother and/or through interaction with various environmental factors (Rodriguez et al., 2015). Initially, the microbial cell count was estimated to exceed the

number of cells of the host by a ratio of 10:1 (Savage, 1977). However, this has been recently recalculated, and it was found that the ratio between microbial cell count and host cell count is closer to 1:1. Based on the current estimates, the microbial cell count per individual is expected to range from 10^3 in the stomach to 10^{11} in the colon, resulting in approximately $3.8*10^{13}$ total microbial cells (Sender et al., 2016). The gut microbiota is the most populated community of microbes in the human body. It is estimated that the number of unique species in the gut is between 500 – 1000 (Turnbaugh et al., 2007) and harbours approximately 2 million bacterial genes (Gilbert et al., 2018). Despite the large population of microbiota in the gut, a single person is expected to have approximately 160 species on average (Qin et al., 2010). It is interesting to note that various taxa can carry out the same metabolic functions and therefore, despite considerable variations in taxonomic diversity across individuals, the microbiota can have a level of functional redundancy. Hence, investigating the relationship between taxonomic and functional diversity can reveal insights into the mechanistic role of gut microbiota (Eng and Borenstein, 2018).

Human microbiota research has documented that the microbiota carries out a variety of functions in the gut such as degradation of plant polysaccharides and production of vitamins, essential amino acids and SCFAs that are beneficial for the host (Rowland et al., 2018). The gut microbiota is also partly responsible for developing and maintaining immune homeostasis of the host by interacting with the host through the intestinal mucosal barrier (Belkaid and Hand, 2014). The mucosal barrier and the intestinal epithelial cells are the primary component of the intestinal mucosal barrier. It serves as the primary interface for interaction between the immune-cells and the gut microbiota (Turner, 2009).

In 1900, Henry Tissier studied the gut microbiota in infants and this is the earliest known research on microbiota (Tissier, 1900). Later in 1958, Eiseman *et al.* showed that FMT treatment could be used to treat pseudomembranous enterocolitis (Eiseman et al., 1958). Subsequently, others have shown that probiotic treatment with *Lactobacillus rhamnosus* is effective against *Clostridium difficile* infection (Schwan et al., 1983). These early studies gave rise to future research on FMT and the role of microbiota in humans. Germ-free mice/rat models showed impaired development, physical and physiological alterations compared to normal mice/rat models which highlighted the role of microbiota in host development (Gordon and Bruckner-Kardoss, 1961;Abrams et al., 1963;Sharma et al., 1995;Petersson et al., 2011). Research on acquisition and colonisation of microbiota in the early stage of life provided early insights into the host-microbiota relationship (Yoshioka et al., 1983;Tamburini et al., 2016). It highlighted the immune-modulatory effect of microbiota in the host and the beneficial capability of the gut microbiota in immune homeostasis (Sudo et al., 1997;Rakoff-Nahoum et al., 2004;O'Mahony et al., 2008;Smith et al., 2013).

The gut microbiota community is primarily investigated by analysing faecal samples from the host. Microbiota communities have also been isolated and studied from other body sites such as oral, nasal, skin, and biopsy samples. The 16S rRNA gene is a molecular marker that is used for phylogenetic studies as it is conserved across bacteria and archaea. It is extensively used for identification and classification of unculturable microbes.

Characterising the microbiota composition from various body sites of multiple individuals has shown that the microbiota across the same body site is likely to have similar profiles whereas the microbial profile can be dissimilar between different sites

(Human Microbiome Project, 2012). This is within reasonable assumption that the individuals are representative of healthy population with similar dietary habits and other host factors like similar age range, no antibiotic medications etc. The microbiota population changes throughout the lifespan with distinct patterns at different stages (Nagpal et al., 2018). Studies have reported that various immune-associated, metabolic, psychological, cancer and cardiovascular diseases are associated with undesirable changes in the microbiota community, potentially due to its role in degrading food materials, metabolite production and interaction with the immune system (Belizario and Faintuch, 2018). The last 10 years of microbiota research have explored the various roles of gut microbiota, their stability in the human host, individual differences across hosts, functional potential, environmental and dietary effects, and their role in regulation of the host immune system (Human Microbiome Project, 2012;Schnorr et al., 2014;O'Keefe et al., 2015;Falony et al., 2016;Zhernakova et al., 2016;Das et al., 2018). Based on large, multi-cohort studies of the microbiota profiles, various factors like intra- and inter-individual differences, genetics, diet, medications, anthropometric measures, ethnicity, age, socio-economic factors and interactions with the environment have been identified as influencing the gut microbiota structure (Arumugam et al., 2011;Goodrich et al., 2014;Obregon-Tito et al., 2015;Falony et al., 2016;Zhernakova et al., 2016;Rothschild et al., 2018;Johnson et al., 2019;Pasolli et al., 2019). This chapter summarises the existing evidences on the interaction between the gut microbiota and some of the prominent factors like diet, medications, and environment. Furthermore, the pathogenesis of various immune disorders and the association with the gut microbiota is also presented. Finally, various bioinformatics approaches developed to study the gut microbiota are discussed.

2.1 Microbiota and the environment

The colonisation by the gut microbial community in the host is considered to begin pre-birth but recently it has been shown that there is no placental microbiota and the detected species are likely to be pathogens or contaminants (de Goffau et al., 2019). However, rapid acquisition, development and stabilisation of the microbiota community occurs after birth and continues for the first few years (Mackie et al., 1999;Fanaro et al., 2003;Tamburini et al., 2016). Various key factors are considered to influence this acquisition and development of gut microbiota in the host. Among them, transmission from mother to infant is prominent as early microbial colonisation in the infants were found to be maternally derived e.g. milk, vaginal, mouth and skin. The acquisition and colonisation of the maternally derived microbiota occurs during the gestation period and birth through vertical transmission (Palmer et al., 2007;Dominguez-Bello et al., 2010;Bokulich et al., 2016;Ferretti et al., 2018). The vaginally delivered infants had a similar microbiota as their own mother's vaginal microbiota whereas the caesarean section infant's microbiota resembled the mother's skin microbiota. Interestingly the gut microbiota profiles in infants born through caesarean section had a decreased taxonomic diversity compared to vaginally delivered babies (Bokulich et al., 2016).

Interaction with pets alters the microbiota of infants with increased abundances of *Ruminococcus* and *Oscillospira* (Tun et al., 2017). These taxa are negatively linked with development of allergy and obesity and might be associated with decreased risk in infants. The abundance of family *Streptococcaceae* was lower upon exposure to pets that may result in reduced risk of *Streptococcal* infection.

Exposure to a rural or urban environment also contributes to the acquisition of the gut microbial community. Urban dwellers were observed to have a reduced

microbiota diversity which were associated with changes in their lifestyle, antibiotics usage, standard of hygiene and isolation from general flora and fauna (Popkin, 1999;Turner et al., 2004;Sjogren et al., 2009). The hunter-gatherer's population from Tanzania, rural healthy individuals from Papua New Guinea and Amerindians have a higher species richness compared to urban western population (Schnorr et al., 2014;Clemente et al., 2015;Martinez et al., 2015). A recent study by Vangay *et al.* found that the diversity in the gut microbiota composition was reduced after a period of acclimatisation among the non-native people coming to USA (Vangay et al., 2018). The native microbial species that were observed in the non-USA population were lost. The group also observed that strains belonging to the genus *Prevotella* were replaced by *Bacteroides* strains. Social interactions with others and cohabitation have also been associated with increased richness and shared taxa among spouses (Dill-McFarland et al., 2019).

The loss of diversity in the microbiota has been associated with decreased immune tolerance and various chronic inflammatory conditions in urban populations such as allergic reactions and asthma (Garn and Renz, 2007;Haahtela et al., 2013). The "hygiene hypothesis" is a concept which suggests that children exposed to various microbes in early childhood experience proper development of the immune system, which in turn results in effective immune tolerance (Strachan, 1989;Garn and Renz, 2007). Studies comparing rural and urban children have shown that rural children have reduced incidences of asthma, which in part is attributed to the higher microbial diversity in the rural environment (Ege et al., 2011). Exposure to an environment with high microbial diversity can be a source of various beneficial or commensal bacteria, which could reduce the risk of inflammatory diseases (Rook et al., 2013;Zhou et al., 2016).

2.2 Microbiota and diet

An infant's diet is another major determinant of microbiota colonisation. A difference was observed in the taxonomic composition of the gut microbiota of breast-fed infants compared to formula-fed infants (Guaraldi and Salvatori, 2012). The microbial population is more stable amongst the breast-fed babies compared to formula-fed infants, but the differences disappear after 1 year of life with introduction to solid foods (Stark and Lee, 1982;Mackie et al., 1999;Adlerberth and Wold, 2009). *Bifidobacterium infantis*, *Staphylococcus epidermidis* and *Lactobacillus rhamnosus* were identified to be dominant in breast-fed infants whereas *Bifidobacterium fragilis*, *Clostridium* and *Streptococcus* species were linked with formula-fed infants (Mackie et al., 1999;Penders et al., 2006;Adlerberth and Wold, 2009). *Bacteroides*, *Clostridium*, *Enterococci* and *Streptococci* genera emerged with introduction of solid food diet (Stark and Lee, 1982;Mackie et al., 1999;Adlerberth and Wold, 2009).

The diet of an individual is the primary source of nutrients for the gut microbiota and a person's diet reflects gut microbiota composition and community structure (Claesson et al., 2012;Zmora et al., 2019). One of the key roles of the microbiota is the breakdown of non-digestible food products such as complex polysaccharides, and biosynthesis of vitamins and essential amino acids (Salyers et al., 1978;Rowland et al., 2018). Various metabolites produced by the microbiota in the intestine can serve as necessary substrate for other microbes whereas other metabolites can have a direct or indirect impact on human health. Bacterial metabolites like SCFAs, secondary bile acids and TMA are known be associated with human health outcomes (Vinolo et al., 2011;Janeiro et al., 2018;Nguyen et al., 2018).

SCFAs can modulate the immune system and can confer protective effects. An increased consumption of dietary fibre is linked with reduced inflammation due to

breakdown of dietary fibres into SCFAs by the gut microbiota (Bohmig et al., 1997;Andoh et al., 1999). Butyrate, propionate, and acetate are the major SCFAs produced by the bacteria. SCFAs serve as a major source of energy for the epithelial colonic cells and the bacterial population. Catabolism of SCFAs by the epithelial cells occurs through β-oxidation that also results in maintenance of anaerobic conditions in the intestine (Byndloss et al., 2017). Anti-carcinogenic activities of SCFAs have also been observed against colon cancer cells by inducing apoptosis and preventing proliferation of the cancer cells (Bindels et al., 2012). They also induce mucin production, which is necessary for maintaining the integrity of epithelial barrier function (Willemsen et al., 2003). Interaction of SCFAs with the liver is associated with regulation of lipogenesis, fatty acid synthesis, cholesterol metabolism, and appetite (Frost et al., 2014). Various species of bacteria are known to be producers of different types of SCFAs. *Faecalibacterium prausnitzii*, *Ruminococcus bromii*, *Clostridium leptum*, *Eubacterium rectale* and *Roseburia* species are known butyrate producers (Louis and Flint, 2017). Pathways necessary for propionate formation are found in *Bacteroides fragilis*, *Megasphaera elsdenii*, *Clostridium neopropionicum*, *Propionibacteria* species and *Veillonella* species (Hosseini et al., 2011). Known producers of acetate are *Blautia hydrogenotrophica*, *Ruminococcus bromii*, *Bifidobacterium* species and *Akkermansia muciniphila*, which can produce both acetate and propionate (Derrien et al., 2004).

Primary bile acids produced by the liver and secreted into the small intestine are metabolised by the gut microbiota into secondary bile acids. These secondary bile acids serve as signalling molecules for various host-associated pathways that affect the host metabolism (Wahlstrom et al., 2016). Deconjugation of the primary bile acids is carried out by BSH enzymes and this enzyme is found in all major bacterial divisions

(Wahlstrom et al., 2016). Formation of a diverse variety of secondary bile acids is carried out by species belonging to *Bacteroides*, *Bifidobacteria*, *Clostridium*, *Eggerthella*, *Escherichia*, *Eubacterium*, *Lactobacillus*, *Peptostreptococcus* and *Ruminococcus* genera (Staley et al., 2017). Secondary bile acids can also interact with immune associated host functions directly or indirectly thereby affecting the gut microbial composition. Various diseases such as IBD, metabolic syndrome, and CRC have been associated with an altered bile acid profile and composition (Jia et al., 2018). Individuals diagnosed with CRC who also consume diets containing high levels of saturated fats and red meat show elevated levels of secondary bile acids (Nagengast et al., 1995;Bernstein et al., 2005;Ridlon et al., 2014). Microbial metabolism of nutrients like choline and carnitine generates TMA, which is converted to TMAO in the liver. TMAO is associated with adverse cardiovascular conditions like atherosclerosis (Tang et al., 2013).

The microbiota also produces various other nutrients such as certain vitamins and amino acids. Vitamin B and K are synthesised by species of *Bacteroides* and *Bifidobacteria*, *Escherichia coli*, *Bacillus subtilis*, and lactic acid bacteria (LeBlanc et al., 2013;Rowland et al., 2018). Lysine is produced by the gut microbiota and it can contribute up to 20% of the lysine measured in the circulatory system of the body (Metges, 2000). Other amino acids produced by the microbiota can also contribute to the maintenance of the amino acid pool in the host.

The use of food additives and sweeteners can disrupt the composition of the gut microbiota. An increase in the abundance of pro-inflammatory bacteria was observed in mouse models consuming emulsifiers or sweeteners (Chassaing et al., 2015;Nettleton et al., 2016;Bian et al., 2017). Other dietary habits like a gluten-free diet, vegan diet or a diet low in FODMAP have been linked with differences in the

microbial metabolite profiles and gut microbiota composition (McIntosh et al., 2017). An *in-vitro* study by Bevilacqua *et al.* found that gluten-free bread affected the gut microbiota from coeliac individuals in a positive manner by promoting *Lactobacillus* and *Bifidobacterium* species growth (Bevilacqua et al., 2016;Lebwohl et al., 2017). Consumption of dietary supplements like probiotics and prebiotics can promote the health of individuals. Probiotics include live microbes, such as species of *Bifidobacterium* or *Lactobacillus*, which can produce key bioactive metabolites or can interact with the immune system to confer benefits to the host. E.g. restoring immune homeostasis by regulating T-reg cells, increased gastrointestinal barrier function, and inhibiting pathogen growth (Azad et al., 2018). Prebiotics are dietary substrates consumed by an individual and metabolised by the gut microbiota providing health benefits like generation of SCFAs. Dietary fibres and MACs are considered as prebiotics. However, the definition of prebiotics is ill-defined and the concept is currently debated (Bindels et al., 2015).

The microbiota is associated with obesity in a diet dependent manner. Introduction of microbiota from obese individuals into germ-free animals resulted in more weight gain compared to mice receiving microbiota from normal weight individuals (Goodrich et al., 2014). Introduction of taxa such as *Christensenella* into germ-free mice correlated with reduced body fat and lack of weight gain. *Oscillospira* and *Akkermansia muciniphila* were found to be negatively associated with fat and obesity in human studies (Everard et al., 2013;Goodrich et al., 2014;Le Roy et al., 2018). Evidence suggests that lower microbial diversity is linked with long-term weight gain and is also reflective of other inflammatory conditions like IBD and PSA (Manichanh et al., 2006;Scher et al., 2015;Menni et al., 2017). A study on the elderly Irish population has shown that habitual diet is associated with microbiota profile and

the microbiota community structure and habitation reflected the frailty and health status of the cohort (Claesson et al., 2012). This shows that a variety of inter-related mechanisms link diet, immune modulation, and microbiota composition within the host.

2.3 Microbiota and medication

Various drugs, medications and supplements consumed by an individual have an effect on the GIT and the gut microbiota. Consumption of antibiotics has a clear effect on the gut microbiota and is used extensively in animal studies to obtain microbiota free animals as opposed to working with germ-free animal models, which can be more expensive and laborious (Kennedy et al., 2018). Use of broad-spectrum antibiotics can wipe out mutualistic and commensal microbes thereby negatively affecting the health of the host due to depletion of SCFA producers, lack of immune modulation and opportunity for growth of pathobionts. Despite the depletion of the gut microbiota due to antibiotic usage, the composition of the gut microbiota tends to restore itself to the pre-treatment composition after completion of short-term use of antibiotics (Palleja et al., 2018). However, the loss of diversity due to repeated antibiotic use can result in an altered gut microbiota composition that promotes the growth of pathogens. *Clostridium difficile* infection, which is a hospital-acquired infection responsible for antibiotic-associated diarrhoea, is a common example of infection due to susceptibility caused by antibiotic treatment and a disrupted microbiota (Jernberg et al., 2010). Introduction of diverse group of bacteria (FMT) from healthy individuals into a CDI mouse model have been effective in reducing experimental *C. difficile* infection (Lawley et al., 2012). Overall, it is important to account for the use of antibiotics in human observational and clinical trial microbiota studies.

Two large European observational studies primarily comprised of people from the Netherlands and Belgium have shown the effect of various medications and other covariates (lifestyle, diet, heath, anthropometrics, blood parameters, bowel information and sex) on microbiota composition (Falony et al., 2016;Zhernakova et al., 2016). Individual factors like BMI, stool consistency, age, sex and diseases are associated with gut microbiota diversity but medication explained the most variation in the microbiota composition. Amongst various medications, the gut microbiota composition is significantly associated with various over-the-counter-drugs and prescribed non-antibiotics drugs. These drugs include but are not limited to, PPIs, metformin, and psychotropic medications as well as NSAIDs and vitamin and mineral supplements, which are also reported by other groups (Jackson et al., 2016;de la Cuesta-Zuluaga et al., 2017;Otani et al., 2017;Skrypnik and Suliburska, 2018;Cussotto et al., 2019). The diversity of the microbiota correlated with comorbidity and polypharmacy in the elderly individuals also (Ticinesi et al., 2017). This highlights the necessity to understand the impact of multiple drug treatments and the presence of multiple disorders/conditions on the microbiota of patients. Drug-induced microbiota alterations can result in higher susceptibility to infection.

Various studies report that long-term use of PPIs is associated with creating a favourable environment for enteric pathogens like *Helicobacter pylori* and *Clostridium difficile* (Hagiwara et al., 2015;Trifan et al., 2017). Clooney *et al.* show a significant difference in the Firmicutes to Bacteroidetes ratio between PPI and non-PPI users (Clooney et al., 2016). Jackson *et al.* observed increased abundance in Streptococcaceae and Micrococcaceae among PPI users in Twin UK samples (Jackson et al., 2018) which were also previously identified in other independent studies. Frequent use of NSAIDs causes ulcers and erodes the mucosal lining in the small

intestine. In germ-free and gnotobiotic rat models exposed to *Escherichia coli* or *Eubacterium limosum*, NSAIDs caused ulceration in the small intestine. However, the unexposed germ-free rat models were resistant to NSAID induced damage or if *Bifidobacterium* or *Lactobacillus* were present (Uejima et al., 1996). This suggests that certain bacteria may aggravate NSAID-induced damage while others may confer protection. Otani *et al.* reported that small intestinal bacterial overgrowth, which is a condition characterised by excessive bacterial population in the small intestine, is a risk factor in NSAID-induced damage to the small intestine (Otani et al., 2017). PPIs intensified NSAID-induced damage to the small intestine according to a study by Wallace *et al.* (Wallace et al., 2011). Studies have also revealed that metformin, which is a widely used medication for type II diabetes, can partially restore the altered gut microbiota composition and increase the abundance of SCFA-producing bacteria and mucin-degrading *Akkermansia muciniphila* in diseased individuals (Forslund et al., 2015;de la Cuesta-Zuluaga et al., 2017). Vitamin D is a micronutrient commonly supplemented with calcium and is necessary for bone health. Vitamin D has an immunomodulatory role and decreased levels are associated with an increased inflammatory state (Cantorna, 2006). Deficiency of vitamin D and polymorphisms in the VDR are both associated with an altered gut microbial composition (Jin et al., 2015;Luthold et al., 2017).

Maier *et al.* did an extensive analysis on the effect of over 1000 non-antibiotic drugs on 40 gut bacterial strains and observed antimicrobial properties in 24% of the drugs that inhibited the growth of at least one bacterial strain (Maier et al., 2018). Various antipsychotics represented the largest drug class. The molecular mechanism of interaction for most of these drugs remains unknown. Another recent study by Zimmermann *et al.* highlighted the metabolic capability of at least one strain from 76

gut bacteria on 65% of oral drugs out of 271 in mice models (Zimmermann et al., 2019). Zimmermann *et al.* observed that the bacterial strains could metabolise between 11 and 95 different drugs. Grouping these strains based on their drug-metabolism capacity resembled their phylogenetic relationship at phylum-level. These highlight the effect of medications on gut microbiota and vice-versa which may explain varying drug responses, generation of functionally distinct or toxic metabolites or masking the disease signature associated with the gut microbiota.

2.4 Microbiota and the immune system

The various functions carried out by the gut microbiota result in direct or indirect interaction with the host immune system through the large intestine. Apart from absorbing minerals and water from undigested waste and excretion, the large intestine is also responsible for providing protection against harmful microbes, immune stressors and maintaining homeostasis for the native microbiota. This is achieved through proper functioning of various mechanism in the large intestine. These mechanisms include presence of the mucosal layer, generation of antibodies and AMPs, and activation of immune cells. The epithelial layer is a single columnar cell type layer present in the inner lining of the intestinal cavity, which acts as a physical barrier. Its purpose is absorption of nutrients and secretion of the mucous. Epithelial cells connect and communicate with a rich population of immune cells from the lamina propria and lymphoid tissue regions (Goto and Kiyono, 2012)

The stability of the gut microbiota and the integrity of the barrier is necessary for maintaining immune homeostasis. Disruption of the epithelial barrier due to erosion of the mucosal lining or other damage can alter the permeability and allow diffusion of antigens or bacteria into host tissues, resulting in inflammation (Yu,

2018). Pathobionts may also spread in an individual's body harming the host by promoting inflammation, colitis, ulcer, or cancer (colon, gastric). The importance of gut microbiota is revealed in studies with germ-free animal models that show improper maturation of the immune system, nutritional deficiency, poor growth and development, increased intestinal permeability and an altered neurochemical profile (Wostmann, 1981). Functional immaturity of the immune system can be observed in infants where the gut microbiota is not established fully (Adlerberth and Wold, 2009).

The immune system function decreases as age increases, which is also associated with persistent low levels of inflammation. This concept is called inflammaging due to increased levels of pro-inflammatory cytokines observed in the serum along with declining numbers of naive $CD8^+$ T-cells and increased memory $CD8^+$ T-cells (Franceschi et al., 2007;Czesnikiewicz-Guzik et al., 2008). Similar effects have been observed in B-cells which shows a decline in B-cells generation and increased number of memory B-cells and intrinsic defects in B-cells (Scholz et al., 2013).

In the intestine, the T-reg cells are responsible for modulating inflammation due to microbiota- and dietary-mediated factors. In contrast, CD4+ effector T-cells contribute to increased intestinal inflammation (Lee and Kim, 2017;Whibley et al., 2019). T-reg cells regulates Th cells and along with IgA is responsible for maintaining a stable and diverse gut microbiota, preserving commensal tolerance (Powrie et al., 1993;Cong et al., 2009;Kim et al., 2016). Regulation and expression of Th17 RORγt transcription factor by T-reg cells is dependent on the gut microbiota whereas RORγt-T-reg cells are associated with dietary immune modulator molecules (Sefik et al., 2015;Kim et al., 2016). T-reg cells expressing GATA3 confer protection against chronic inflammation and stimulate T-reg cell accumulation in the intestine. The T-

reg cells are also known to be involved in tissue repair and epithelial stem cell renewal (Yu et al., 2015;Zhang et al., 2017;Biton et al., 2018). Studies on mice have shown that increased susceptibility to gut pathogens is associated with T-reg cell depletion (Wang et al., 2014).

Various taxa from different genera namely *Clostridium*, *Lactobacillus*, *Bifidobacterium*, *Bacteroides*, *Ruminococcus*, *Faecalibacterium*, *Roseburia* and *Helicobacter* are associated with various cytokines and T-cells (Schirmer et al., 2016;Chai et al., 2017). Evidence from mice and human studies have shown that SCFAs can promote T-reg cell numbers and are linked with a decrease in inflammation (Maslowski et al., 2009;Arpaia et al., 2013). Immunosuppressive activity of SCFAs have also been observed with APCs and DCs (Bohmig et al., 1997;Nastasi et al., 2015;Park et al., 2019). In-vivo studies have shown a reduction in IgE-associated eczema in infants treated with probiotics and prebiotics (Abrahamsson et al., 2007). Similarly, other oligosaccharides can increase IgA and anti-inflammatory IL-13 levels and decrease pro-inflammatory cytokines like IL-1β. Metabolism of protein-based dietary products generates metabolites which are generally associated with stimulation of colonic inflammation (Lecerf et al., 2012). Dietary supplementation of amino acids to athletics and individuals performing intense exercises have been linked with decrease in levels of IFN-γ, IL-10, increase in levels of IL-1β, C-reactive proteins and other WBCs like lymphocytes and neutrophils (Murakami et al., 2009;Kraemer et al., 2014). Metabolism of aromatic amino acids by the gut microbiota generates p-cresol and its derivative p-cresylsulphate which contributes to pro-inflammation and supresses activation of monocytes and lymphocytes (Schepers et al., 2007). Isomers of conjugated Linolenic acid influence T-cell activation and levels of IL-1β, IFN-γ, IL-10 and IL-5 in a dose dependent manner (Tricon et al., 2004;Song et al., 2005).

Interaction and association of various host cells in the large intestine with the gut microbiota is illustrated in **Figure 1** (obtained with permission).

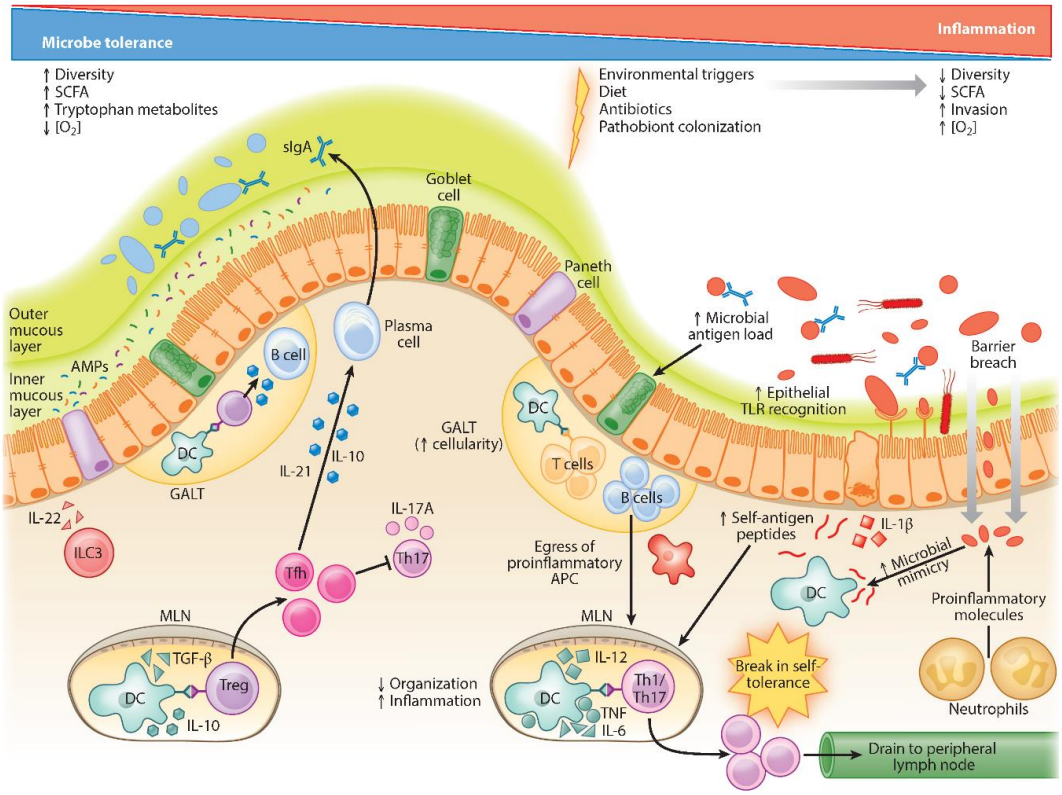


Figure 1, which is obtained from Brown *et al.* (Brown et al., 2019), depicts the various microbiota-associated features, such as diversity, metabolites, diet, medication, other environmental factors and their interaction with the gut intestinal barrier cells and the immune cells. The activation and mechanism of cytokine-mediated immune homeostasis is also depicted. The depletion of the mucosal layer allows greater interaction between the microbes or microbial metabolites and the immune cells.

2.5 Microbiota and diseases

The various gut microbiota associated factors discussed until now highlight the complex interactions between the gut microbiota and the host. The modulation of any of these factors can induce changes in the community structure and the composition of the gut microbiota. Understanding the clinical implications of how the gut

microbiota influences human health is necessary to uncover their roles in various diseases. However, various confounding factors with the potential to affect the microbiota must be accounted for while investigating microbiota-disease association. This is necessary to reduce false positives and to ensure more robust observations. Although recent studies have identified various microbiota association with different types of diseases, human microbiota research is still at a very early stage. Studies across the world have identified alteration in microbial diversity and community profiles in various diseases including identification of potential specific marker taxa associated with health and/or disease (Kho and Lal, 2018). Certain key taxa identified were associated with more than one disease (Duvallet et al., 2017). These microbial alterations are associated with changes in immune associated conditions like IBD, Coeliac disease, RA, AS, PSA, MS, allergy and Diabetes type I (Clemente et al., 2018;Han et al., 2018;Pascal et al., 2018), metabolic and cardiovascular conditions like obesity, Diabetes type II, hypertension and atherosclerosis (Jie et al., 2017;Pascale et al., 2018), infectious diseases like CDI (Stecher and Hardt, 2008), neurodevelopmental conditions which includes depression, ASD and Parkinson's disease (Warner, 2019) and CRC (Wu et al., 2009). Many of these studies focus on correlations between the microbiota and the health and disease, lacking insight into defined functional mechanisms. Alternatively, studies with probiotics and FMT show some success in attenuating or achieving remission in certain diseases (Liu et al., 2018;Mullish et al., 2018). However, these reports require further investigation and rigorous research. The evidence suggests a complex connection between microbiota and development of various diseases as opposed to a unidirectional cause-and-effect association. This opens up the potential of gut microbiota not only as a diagnostic measure but also for therapeutic purposes.

3. Immune-associated conditions and the microbiota

The aetiology of various immune-associated disorders is multi-factorial. Epidemiologically various immune diseases are more prevalent in western countries compared to Asian and African demographics. The age of onset also varies for these diseases, though some of them can occur at a very young age. **Table 1** summarises the epidemiology of various immune-associated diseases.

Table 1 Epidemiology of various immune-associated disorders

Diseases	Prevalence (%)				Sex (F:M)	Age of onset (years)	Comments	References
	Europe	America	Africa	Asia				
Osteoporosis	1-15	3-16	No consensus	4-38	8:5	≥ 50	1 in 5 men and 1 in 3 women is at risk. In men, hip fractures occur 10 years later than in women. Northern Europe has more fractures than southern Europe	(Johnell and Kanis, 2006;Wade et al., 2014;Eastell et al., 2016)
Rheumatoid arthritis (RA)	0.33-0.8	0.5	0.1-2.5	0.2-0.3	4:1	≥ 25	Highest in Native American (~5%). High in northern Europe compared to southem region. Elderly onset RA (EORA) ~65 years	(Peschken and Esdaile, 1999;Silman and Pearson, 2002;Sokka et al., 2009;Tobon et al., 2010;Usenbo et al., 2015;Kobak and Bes, 2018)
Ankylosing spondylitis (AS)	0.15-2.2	0.1-0.32	0.07	0.009-0.78	1:3	≥ 25	Low prevalence in Finland. higher prevalence in Norway	(Gran et al., 1985;Kaipiainen-Seppanen et al., 1997;Hukuda et al., 2001;Feldtkeller et al., 2003;Liao et al., 2009;Dean et al., 2014)
Psoriatic arthritis (PSA)	0.17-0.42	0.02-0.16	1-4.4	0.001-0.02	0.7:1 – 2.1:1	≥ 30	Amongst psoriasis patients, prevalence of PSA is 6-42%. Reported in kids and elderly also	(Torre Alonso et al., 1991;Gladman et al., 2005;Liu et al., 2014;Usenbo et al., 2015)
Inflammatory bowel diseases (IBD)	0.001-0.51	0.001-0.32	0.01-0.02	0.001-0.06	Varied based on age and disease type	≥ 10	Reduced risk of Crohn in female in childhood. Increased thereafter. No difference in risk based on sex for Ulcerative colitis	(Dabritz et al., 2017;Ng et al., 2018;Shah et al., 2018)
Multiple sclerosis (MS)	0.16	0.15-0.45	0.03-0.06	0.03	3:1	≥ 20	Prevalent in population with European ancestry. Can occur in childhood. Decline in life expectancy	(Kingwell et al., 2013;Collaborators, 2019;Wallin et al., 2019)

Prevalence in Europe includes all regions of Europe, not only Western Europe. Prevalence in America includes both North and South America. Prevalence in Asia excludes the Middle East.

The primary basis for these disorders is the aberrant functioning of the immune system that results in increased production of pro-inflammatory cytokines and influx of active immune cells into various sites depending on the disease (Chen et al., 2018;Clemente et al., 2018). Despite host genetic variation and heritability playing a major role in the pathogenesis of various immune diseases, there is growing evidence to suggest that external factors such as environment, diet, and the gut microbiota are also involved (Vojdani et al., 2014). It is established that the gut microbiota is necessary for proper functioning of the immune system and various studies have highlighted its potential role in the pathogenesis/onset of various conditions like arthritis, MS, IBD, and osteoporosis (Clemente et al., 2018;Jones et al., 2018). A recent study by Forbes *et al.* investigated IBD, MS and RA for a common microbiota signature and observed that microbial diversity was lower in all cases compared to healthy individuals (Forbes et al., 2018). Species belonging to the genera *Actinomyces*, *Eggerthella*, *Clostridium* III, *Streptococcus* and *Faecalibacterium* were more abundant in all the disease samples. The study also identified various individual taxa unique to each disease and showed that the identified taxa were able to predict CD most accurately, followed by MS, RA, and UC. The microbial alterations were consistent with other studies carried out on individual immune associated disorders. Clemente *et al.* reviewed and summarised the current understanding of the relationship between microbiota, microbial metabolites, immune cells, cytokines and immune-associated diseases (Clemente et al., 2018) which is highlighted in **Figure 2** (obtained with permission).

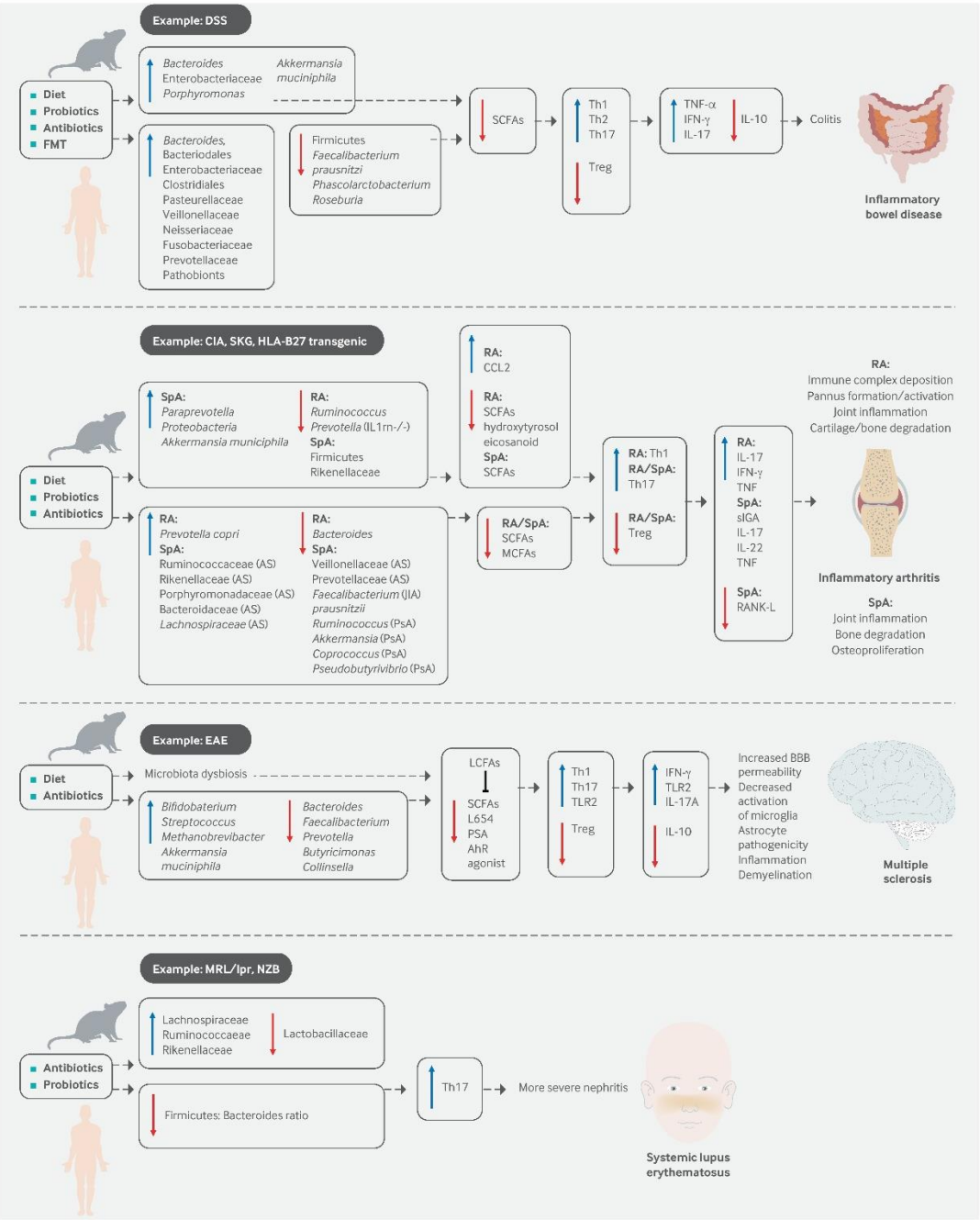


Figure 2 obtained from the review by Clemente *et al.* (Clemente et al., 2018) highlights the increase/decrease of various taxa, metabolites, immune cells, and cytokines in relation to different immune-associated diseases (IBD, RA, SpA, and MS). It also highlights the results from mice studies and human studies, including various treatments and dietary information.

3.1 Osteoporosis

Osteoporosis is a systemic skeletal disorder characterised by weakening of the bone due to deterioration of the bone tissue and low BMD that results in increased susceptibility to fractures and bone frailty (Hernlund et al., 2013). Most osteoporotic fractures occur in the hips, wrist, and spine. Osteoporosis is a commonly observed age-related condition in the elderly population and it is most common in post-menopausal women. Osteoporosis does not manifest any symptoms and hence it is not usually detected until fractures occurs due to minor stress. However, not all fractures in old age are related to osteoporosis (Eastell et al., 2016). Osteopenia is the precursor state of osteoporosis where BMD is reduced but not as dramatically as in osteoporosis. The clinically determined BMD score determines the osteoporosis status. Osteoporotic individuals are defined as having BMD values below -2.5 standard deviations from the mean of peak/healthy bone mass of adult healthy individuals whereas people diagnosed with osteopenia have BMD values below -1.5 standard deviations. The standard deviation measures are defined as T-scores (Kanis et al., 1994;Kanis et al., 2013). Genetics plays a major role in the aetiology of osteoporosis, however other factors such as smoking, exercise or manual labour have also been associated (Eastell et al., 2016).

Currently, osteoporosis is detected by measuring BMD through DEXA scans and is a treatable condition if diagnosed early (Eastell, 1998). Untreated or undiagnosed cases can lead to high fracture rates, loss of mobility, and affect the quality of life. The diagnosis of osteoporosis is based on quantifying BMD determined by DEXA results, as it is considered as the best predictor for the risk of fracture. FRAX is another diagnostic measure that is used to evaluate the probability of risk of fracture

up to 10 years (Kanis et al., 2008). Serum levels of procollagen type I amino-terminal propetide and serum C-terminal telopeptide of type I collagen are two recommended markers for bone formation and resorption respectively which allow characterisation of the disease and the response towards treatment (Vasikaran et al., 2011;Diez-Perez et al., 2012).

The onset of osteoporosis is multifactorial but the primary contributors are genetics (heritability accounts for 50-70% of cases), and sex hormones, especially oestrogen levels in postmenopausal women (Szulc et al., 2006;Hernandez-de Sosa et al., 2014). In men, deficiency of androgen leads to sarcopenia, resulting in loss of body strength and increased probability of falling and fractures (Wiren et al., 2012). Bone mass is regulated by various key genes like vitamin D receptor, TGFβ, collagen type I-α, LRP5, LRP6, and other genes from the Wnt pathway. Variations in these key genes have been associated with alterations in bone mass (Ralston et al., 2006;Uitterlinden et al., 2006;Langdahl et al., 2008;van Meurs et al., 2008). Assessment of individuals over the age of 50 should be carried out based on risk factors like smoking status, genetics, fractures, low BMI, sarcopenia, and frailty status. Despite having the same T-scores, without at least one risk factors, the risk of fractures in postmenopausal women is low compared to women with risk factors (Siris et al., 2011).

Osteoblasts, osteoclasts, and osteocytes are the specialised cells responsible for maintaining the balance between bone resorption and bone formation. Although the immune system is involved in activation of osteoblast and osteoclast cells, the role of autoantibodies has been identified only when other autoimmune diseases are involved (Iseme et al., 2017). Osteoblasts, and osteoclasts have high affinity oestrogen

receptors and their interaction with various ligands triggers various effects like activation or inhibition of IL-1, IL-6, IGF1, TNF, and TGFβ (Weitzmann and Pacifici, 2006). Activated T-cells are also the source of RANKL and TNF-α which are responsible for destruction of bone during inflammatory conditions. Studies with animal models lacking T-cells have shown normal or higher BMD. However, the exact links are not fully understood. The secretion of various cytokines by T-cells can direct or indirectly influence bone modulation (Hofbauer and Schoppet, 2004;Srivastava et al., 2018).

Supplementation of essential minerals and vitamins like calcium and vitamin D in the diet of people at risk is recommended as a preventive measure against the development of the disease. However, studies have shown that increased intake of calcium does not prevent development of the postmenopausal osteoporosis or fracture prevention but reduces age-associated BMD loss (Elders et al., 1994;Shea et al., 2002;Avenell et al., 2014). Vitamin-D deficiency can also increase PTH secretion that can indirectly affect osteoclast cells and bone resorption as PTH is responsible for maintaining serum calcium levels (Khundmiri et al., 2016). Long-term use of bisphosphonate, one of the most widely used agent for treatment of osteoporosis (Lewiecki, 2010), has been reported to cause gastro-intestinal problems, renal failure, and osteonecrosis of the jaw (Vescovi and Nammour, 2010;Khosla et al., 2012). Denosumab is a monoclonal antibody developed to target RANKL inhibiting osteoclast activity showing significant increase in BMD. Denosumab is given as subcutaneous injection every 6 months, however if injection is missed, a rapid relapse is observed (Miller et al., 2008).

Proper maturation of the immune system is partly dependent on the host microbiota. Various studies have shown alteration in bone development in germ-free and antibiotic treated animal models. The lack of microbial community has been associated with an increase in bone mass and strength. In contrast, other animal studies have reported that lack of microbiota can negatively affect bone properties (Pytlik et al., 2004;Sjogren et al., 2012;Nobel et al., 2015;Guss et al., 2017;Guss et al., 2019). During inflammation, gut microbiota can influence the T-cells to produce cytokines responsible for bone mineral resorption (Hsu and Pacifici, 2018). Several studies have been carried out to identify probiotics that confer beneficial effects on bone health. These beneficial effects can include improvements in vitamin and mineral uptake, production of hormones or providing a protective effect against inflammation and bone resorption (Xu et al., 2017).

Very few studies have been published with human cohorts exploring the role of the microbiota in bone homeostasis, especially concerning osteoporosis. Blanton *et al.* explored the gut microbiota in malnourished children and identified a microbiota associated role in physical development and growth which also showed altered bone morphology in the malnourished groups (Blanton et al., 2016). Excluding this study, only two studies have been carried out on an adult human cohort. The first was carried out by Wang *et al.* where they observed significant alterations of α- and β-diversity measures between osteoporotic and healthy groups and identified various key discriminating taxa associated with osteopenia and osteoporosis (Wang et al., 2017). Despite this study being carried out on gender- and age-matched samples, the cohort size was very small and lacked information on other confounders. The study also did not control for false positives in their analysis. The second study was carried out on a larger cohort by Li *et al.* where they identified different taxa that were associated with

BMD measures (Li et al., 2019). In this study, the characterisation of the samples was done by separating the BMD measures into two categories based on the median BMD value: low BMD and high BMD. The study did not observe any significant alteration at both α- and β-diversity measures. Inferred pathways associated with LPS biosynthesis were more abundant in the low BMD group. However, this grouping may not be clinically relevant as the risk values may be different from segregation threshold which is an arbitrary cut-off. These two studies also observed contrasting results with respect to the microbiota diversity and its association with bone health. Interestingly, the LifeLines-DEEP cohort, which explored the gut microbiota profiles from a Dutch cohort, observed that osteoporosis showed a small but significant effect-size with a global microbiota profile (Zhernakova et al., 2016). Jackson *et al.* investigated the effect of different diseases and medications on gut microbiota profiles in a UK cohort and showed small but significant associations with fracture risk and calcium medication (Jackson et al., 2018). Bisphosphonate medication, which is routinely used for osteoporosis treatment, has been associated with increased abundance of *Actinomyces* species and bisphosphonate-associated jaw necrosis (Russmueller et al., 2016). Vitamin-D, which is a supplement for promoting bone health, has immune-modulatory properties (Luthold et al., 2017).

These studies highlight the potential role that the gut microbiota plays in bone homeostasis through interactions with the immune system.

3.2 Rheumatic diseases

Rheumatism or rheumatic diseases are disorders of the joints and/or connective tissues characterised by chronic or irregular pain. It is an umbrella term for various disorders

which includes arthritic and non-arthritic conditions. Although the diseases share very little similarity in terms of their pathology and epidemiology, the common symptoms are often long-term pain and stiffness that are difficult to cure. The onset and progress of various arthritic conditions is due to aberrant immune responses resulting in autoimmunity and inflammation. This leads to destruction of the bone, cartilage, and tissues (in more advanced stages). Other symptoms of arthritic conditions include swelling, redness, tenderness, and limited joint movement. Undiagnosed or untreated conditions can severely disrupt quality of life and can result in loss of mobility .

3.2.1 Rheumatoid arthritis

Rheumatoid arthritis (RA) is a chronic, inflammatory disorder known to particularly affect the joints. People with RA have swollen joints with stiffness, pain, and symmetrical inflammation. Common body sites affected include, but are not limited to, smaller joints like hands, feet, and cervical spine as well as knees, shoulders, and other larger joints (Heidari, 2011). It can also result in complications in other parts of the body like lungs, heart, skin, eyes, and degradation of the bone leading to irreversible disability (Scott, 2004;Cojocaru et al., 2010). Autoantibodies are antibodies produced by the immune cells which target and acts on an individual's own proteins or tissues. RA is autoimmune in nature involving production of autoantibodies against citrullinated proteins and IgG. The autoantibodies produced by the human body are RF which target the Fc receptors present in IgG and also ACPAs that interacts with the citrullinated proteins (Malmstrom et al., 2017). However, RA is heterogeneous in nature, involving various genetic and environmental stressors that

can trigger pathogenesis in a variety of ways. Currently there is no established cure for RA, with remission being the most likely outcome from treatments.

RA is classified by serological tests that are related to production levels of autoantibodies and acute-phase reactants (inflammation markers e.g. C-reactive proteins, haptoglobin, plasma fibrinogen) and by investigating the clinical progression of the disease. Currently, there are no diagnosis criteria and the diagnosis of RA is highly individualised. Clinical progression is monitored by studying the joint swelling, morning stiffness and tenderness (Egerer et al., 2009;Aletaha and Smolen, 2018). However, other types of arthritis may also show similar symptoms which makes diagnosis of RA difficult. RA is a systemic condition and is not exclusive to the joints. Manifestation and risk of cardiovascular diseases also increase with RA (Crowson et al., 2013). Currently, the classification criteria recommended by ACR and EULAR, which were defined in 2010, are followed (Aletaha et al., 2010). The three main indexes that are used to measure disease activity are CDAI, SDAI and/or DAS. The remission status of the disease and any improvement is calculated using Boolean remission criteria and ACR improvement criteria (Aletaha and Smolen, 2007;Felson et al., 2011;van Riel and Renskers, 2016).

Development of RA is strongly associated with genetics. However, other factors like sex, obesity, and other environmental factors such as smoking and exposure to silica have also been reported as key risk factors. Heritability of RA was estimated to be approximately 15% in an identical twin study by Silman *et al.* but another study by MacGregor *et al.* found it to be 60% (comprising of both identical and fraternal twins) (Silman et al., 1993;MacGregor et al., 2000). ACPA-positive showed higher risk of RA compared to ACPA-negative individuals (Padyukov et al.,

2011) which may indicate a bigger role of other factors. Human leucocyte antigen (HLA) class II is strongly linked with RA (Gregersen et al., 1987). Similarly, genetic variants of PRL and NFIA are reported in seronegative individuals while AFF3, CD28 and TNFAIP3 gene variants are present in seropositive subjects. Irrespective of serostatus, variation in HLA-DRB1, PTPN22, BLK, ANKRD55 and IL-6ST genes have been associated with RA (Viatte et al., 2016). Smoking and RA are strongly associated, and this association is possibly limited to ACPA-positive patients (Meng et al., 2017). Smoking is linked with increased pro-inflammatory cytokines levels and RA disease activity (Sokolove et al., 2016). Individuals with RA have also shown higher risk of prevalence of cardiovascular conditions which are the most common cause of death among those individuals (Radner et al., 2017).

The onset of RA occurs much earlier than it is clinically detected or classified. Various factors can act independently to contribute to its development by triggering changes like citrullination or other modifications in proteins which can serve as target for generation of autoimmunity but it can also be considered as a part of the normal immune response (Makrygiannakis et al., 2008;van der Woude et al., 2010). The presence of RF and APCAs can be detected a decade before being clinically classified as RA. These markers are also associated with more aggressive disease progression (Houssien et al., 1997;Toes and van der Woude, 2011;Arend and Firestein, 2012).

The synovium is a lining of specialised soft connective tissue in the inner surface of the joint. The synovium serves as the main site for pathogenesis. The synovial barrier is not highly selective and allows leakage of cells and proteins into the synovial fluid. Inflammation in the synovium due to infiltration and presence of CD4+ T-cells, macrophages, and fibroblasts is necessary for disease onset and

progression (Kinne et al., 2000;Klarenbeek et al., 2012;Lefevre et al., 2015). These cells types are also involved in production of pro-inflammatory cytokines and proteases. The influx of various cells into the synovium including mast cells, macrophages, and neutrophils along with FLS results in production of MMPs and collagenases which damage the extracellular matrix proteins, collagens and cartilage (Keyszer et al., 1998;Bluml et al., 2014). Osteoclasts are responsible for breakdown of the bone due to nuclear factor κB produced by T-cells. In the absence of autoantibodies, TNF was shown to increase osteoclast activation resulting in joint damage (Redlich et al., 2002;Schett and Gravallese, 2012). Other studies have highlighted the roles of other cytokines like IL-6, IL-8, IL-1, IL-15, IL-7, IL-17 and interferons that can interact with various receptors (Magyari et al., 2014).

Treatment strategies recommend the use of DMARDs along with short-term dosage of glucocorticoids (Aletaha and Smolen, 2018). Glucocorticoids are a class of steroidal hormones with disease-modifying activity but are not recommended for long-term use due to possible adverse effects (del Rincon et al., 2014). DMARDs block inflammatory processes through both known and unknown targets. Methotrexate is a commonly used synthetic DMARD (Smolen et al., 2017). Other types of DMARDs are targeted synthetic DMARDs that target JAK signalling and biologics (also known as biological DMARDs) which are monoclonal antibodies produced to target specific cell receptors and proteins namely TNF, IL-6 and its receptors, CD80, CD86 and CD20 (Fleischmann et al., 2017;Nam et al., 2017). Combination of both biologics and other synthetic DMARDs can be more effective compared either treatment alone (Smolen et al., 2014) but biologics are costlier and hence, their use as first line of treatment is still debated.

Many studies have explored possible links between RA and the microbiota given that the microbiota plays a key role in immune modulation. In this regard, various animal studies have been carried out with knock-out models of different immune genes to elucidate the role of microbiota in the development of the disease. Rogier *et al.* reported that the microbiota stimulates IL-17 and TLR receptor dependent arthritis in IL-1 receptor negative mice. The IL-1 negative mice also had lower microbiota diversity (Rogier et al., 2017). The same group observed that *Ruminococcus, Parasutterella, Xylanibacter, Barnesiella,* and *Prevotella* were depleted whereas *Helicobacter, Butyricimonas, Rikenella,* and *Streptococcus* were abundant in the IL-1 negative mice. This highlights the role of IL-1 in modulating the microbiota. Germ-free mice show a lack of development of the arthritis phenotype. The development of arthritis was observed in germ-free animals with the introduction of *Lactobacillus bifidus,* other filamentous bacteria or recolonization of microbial flora from collagen induced arthritis susceptible mice (Abdollahi-Roodsaz et al., 2008;Wu et al., 2010;Liu et al., 2016). The SCFAs have an opposite impact on disease severity depending on the knockout mouse type (Mizuno et al., 2017). Investigating the probiotic effect of different bacteria uncovered that *Lactobacillus casei* confers a beneficial effect and suppresses pro-inflammatory cytokine expressions in RA animal models (Mizuno et al., 2017;Pan et al., 2019). Treatment with *Lactobacillus casei* also resulted in increased abundances of *Lactobacillus hominis, Lactobacillus reuteri,* and *Lactobacillus vaginalis* that were decreased in diseased animals before treatment.

A large human study by Zhang *et al.* on a Chinese cohort (212 samples) highlighted the alteration in the gut and oral microbiota profiles of RA patients (Zhang et al., 2015). The study observed that microbiota alteration in the disease state partially reverted back towards normal after treatment of MTX and/or glycosides of the

traditional Chinese medicinal component *Tripterygium wilfordii*. This study also observed that the oral and gut microbiota showed concordance between them and had distinct profiles compared to healthy samples. *Haemophilus* species were found to be depleted in RA patients and negatively correlated with autoantibodies. In patients with onset of RA, increased abundance of *Prevotella copri* was found as well as a decrease in species belonging to *Bacteroides*. Chen *et al.* noted that taxa from the Actinobacteria were more abundant in RA and showed that *Collinsella,* a taxon from the phylum Actinobacteria was associated with increased IL-17A mediated inflammation, gut permeability, and disease severity in humanised mice and human epithelial cell line (Chen et al., 2016). The group also observed that MTX treatment was associated with increased α-diversity suggesting a potential restoration of the gut microbiota. Jeong *et al.* investigated the microbiota profiles of female subjects with RA, which included mainly preclinical and early stages of RA showed that the β-diversity was significantly different compared to healthy cases but not α-diversity (Jeong et al., 2019). The phylum Bacteroidetes was enriched whereas phylum Actinobacteria was decreased in people with RA. Functionality associated with LPS was positively associated with established RA patients and DAS-28 metric. The decrease in Actinobacteria by Jeong *et al.* is in contrast to the findings in the Chen *et al.* study and may be due to difference in patient characteristics, clinical features, and mediation usage. Recently Sultan *et al.* reported that the use of antibiotics increases the risk of RA (Sultan et al., 2019).

A meta-analysis study by Armour *et al.* investigated the microbial functional signature of various disorders that included RA, IBD, colorectal cancer, diabetes II, liver cirrhosis and obesity, and compared these diseases against healthy controls (Armour et al., 2019). The group observed that the functional (protein) richness and

diversity from RA samples was not significantly different from the healthy controls. Only a few functional modules were linked with RA. Modules related to acetate and methane production were increased in RA. The lack of strong association with functionality resulted in lack of sensitivity and specificity for RA classification based on KEGG modules. These human studies consistently reported that both taxonomic and functional richness were very similar to the healthy controls. This may suggest that specific taxonomic alterations may play a prominent role in driving pathogenesis of RA as opposed to reduced/altered diversity which is normally associated with inflammation.

Investigation into the effects of probiotics on RA disease activity have been carried out in small clinical trials. *Lactobacillus casei* was found to attenuate disease activity and increase anti-inflammatory cytokines (Alipour et al., 2014). Zamani *et al.* studied the effect of administering a consortium of bacteria (*Lactobacillus acidophilus*, *Lactobacillus casei* and *Bifidobacterium bifidum*) and observed an improvement in the DAS-28, and reduced levels of CRPs (Zamani et al., 2016). *Prevotella copri* was enriched in new-onset RA individuals who had not undergone any treatment compared to healthy controls. The abundance of *Prevotella* correlated with the decrease in *Bacteroides* and other beneficial microbes suggesting a potential role in development of RA (Scher et al., 2013).

3.2.2 Ankylosing spondylitis

Ankylosing spondylitis (AS) is an asymmetrical, inflammatory disease that affects the axial skeleton. In AS, the bone-tendon interface serves as the site where inflammation originates, leading to both erosion and proliferation of the bone resulting in fusion of

bones (Zhu et al., 2019). AS is a subtype of spondylarthritis (SpA) which is an umbrella term that includes reactive arthritis, spondylitis, and enteritis (Sieper et al., 2015). In 1990s, the ESSG and Amor *et al.* proposed defining the group of SpA as one disease (Amor et al., 1990;Dougados et al., 1991). Based on ESSG criteria, SpA is categorised as axial SpA (axSpA) characterised predominantly by back pain and peripheral SpA characterised by inflammation in the peripheral joints. Later, with inclusion of MRI results, the ASAS developed new criteria. Based on these, axSpA was split into AS (radiographic disease) that show structural damage and bony bridges in the sacroiliac (SI) joints and non-radiographic axial SpA (nr-axSpA) that shows no damage in SI joints (Rudwaleit et al., 2011;Sieper and van der Heijde, 2013). AS is a seronegative condition as there is no RF autoantibodies like those observed in RA. AS has long been associated with the HLA-B27 allele, however the genes IL-17A, IL-23 and ERAP-1 have also been linked with AS (Brewerton et al., 1973;Evans et al., 2011;Taams et al., 2018).

Initially AS was detected based on radiographic evidences, which was essential as it was possible to observe visible sacroiliitis. However, as the disease developed with inflammation, it was not detected in early cases where there was no visible structural damage (Sieper et al., 2002). Currently based on ASAS criteria, classification of AS includes MRI results and a test for HLA-B27 (Rudwaleit et al., 2009b). The patients must also have chronic back pain for at least 3 months at an age of less than 45 years. The sensitivity and specificity of classification is at 82.9% and 84.4% respectively when testing for clinical or laboratory measures like HLA-B27 along with imaging. There are no diagnostic criteria established for AS, hence classification criteria are used commonly for diagnosis (Rudwaleit et al., 2005). The classification criteria reflect the standardised definitions from well-defined cohort of

patients whereas diagnostic criteria require identifying the cause and nature of the disease determined by supportive tests, signs, and symptoms. Diagnosis is carried out by investigating the x-ray reports, investigating inflammation in at least two different sites using MRI, measuring C-reactive proteins levels, testing for HLA-B27, accounting for family history of SpA and response to NSAIDs (van den Berg et al., 2013). A flexible approach is necessary depending on the condition of the patients.

Genetics has been established as the primary risk factor for development for AS. HLA-B27 has the strongest link with AS (Schlosstein et al., 1973;Bowness, 2015). In cases of advanced axSpA, the prevalence of HLA-B27 is estimated to be approximately 80% (Rudwaleit et al., 2009a). The concordance of the risk in monozygotic twins is 63% and in cases of first-degree relatives it is 8.2% (Brown et al., 2000). Genetic variants of ERAP-1 have been associated with increased risk of AS for it is responsible for trimming HLA class 1 peptides, affecting HLA-B27 interactions(Evans et al., 2011). IL-17A and IL-23 are also recognised as important for AS because IL-23 signalling through CD4+ T helper cells causes differentiation of T helper 17 cells that produce IL-17A cytokines (McGeachy et al., 2009;Jethwa and Bowness, 2016). Excluding genetic factors, smoking and hypertension are also recognised as risk factors for AS (Videm et al., 2014).

Onset of AS involves autoinflammation at the specific site due to immune and mechanical stress. The SNPs at IL-23 receptors were also found to be responsible for the promotion of cytokines IL-17F, interferon-γ and TNF (Coffre et al., 2013). The increased expression of IL-17 is related to other cells like neutrophils, mast cells and myeloperoxidase cells (Appel et al., 2011;Noordenbos et al., 2012;Appel et al., 2013). However, their roles in driving AS pathogenesis remain unclear. The overexpression

of IL-23 in experimental models has shown induction of an inflammatory state that resembles that of axSpA and RA (Sherlock et al., 2012). Genetic studies have identified SNPs on the TNF receptor 1 gene that are associated with AS samples (International Genetics of Ankylosing Spondylitis et al., 2013) but the exact role of TNF remains unclear (Braun et al., 1995). In axSpA, both TNF and IL-17 play a major role in cartilage destruction but AS is uniquely distinguished with bone proliferation also (Francois et al., 2000;Herman et al., 2008). Until now, there is no evidence to support the relationship between osteo-proliferation and the inflammation observed in this disease as anti-inflammatory therapy targeting TNF does not inhibit bone formation (Baraliakos et al., 2014). However, it has been suggested that the effect of targeting TNF would take years to show any measurable effect on osteo-proliferation (Haroon et al., 2013).

Use of DMARDs can prevent radiographic damage, but there is limited research showing disease-modifying activity and efficacy on axSpA (van der Horst-Bruinsma et al., 2002;Braun et al., 2006). Use of glucocorticoids is effective at high dosage (Haibel et al., 2014). Primarily, TNF-specific therapy have shown to reduce inflammation, C-reactive protein levels, and improve mobility and functionality of the spine, hence overall improvement in health (Braun et al., 2003;Haroon et al., 2013). Other biologics treatments have been developed that target IL-17, IL-12 and IL-23 which have shown improvements in patients (Baeten et al., 2013;Poddubnyy et al., 2014). Treatment of AS with NSAIDs was effective with partial remission in approximately 10-35% of patients (van der Heijde et al., 2005). Early diagnosis followed by continuous treatment can reduce radiographic progression compared to on-demand treatment (medication only if necessary); however, this still requires

extensive investigation due to risk of side-effects from long-term medications (Wanders et al., 2005).

The role of the microbiota as an environment trigger for disease pathogenesis is being studied in various arthritic conditions. In germ-free conditions, HLA-B27 transgenic rats models do not develop AS but when commensal gut microbiota are introduced, these rats developed AS symptoms and inflammation (Rath et al., 1996;Lin et al., 2014). Studies with germ-free mice also report similar results (Rehakova et al., 2000;Sinkorova et al., 2008). Studies with rat models have identified alterations in bacterial species, namely decreases in Firmicutes and Rikenellaceae. Taxa that are observed to be enriched in rat models are Proteobacteria, *Paraprevotella, Prevotella, Bacteroides vulgatus* and *Akkermensia municiphila* (Lin et al., 2014;Asquith et al., 2016). These alterations have been associated with colonic Th cell expansion and increased expression of Th cell cytokines and IgA antibodies. Asquith *et al.* found that bacterial metabolites like propionate are associated with attenuation of disease (Asquith et al., 2017). The expression of HLA-B27 in AS rat models have been reported to effect microbial metabolites. A very recent study by Gill *et al.* observed that the gut microbiota signature was distinct in different rat models and yet the association with pro-inflammatory cytokines remained. The group identified metabolic pathways related to SCFAs, LPS and steroids that were associated with immune modulation. However, the identified microbial markers did not show any significant difference when compared to wild-type controls (Gill et al., 2019).

Human studies have also shown that up to 30% of IBD patients can develop symptoms of AS (Orchard et al., 2000). The role of gut microbiota in AS pathogenesis is suspected as 61% of AS patients have shown varying degrees of gut inflammation

(De Vos et al., 1989). Analysis of biopsy samples by Costello *et al.* from a small cohort of AS patients and healthy controls revealed that the families Veillonellaceae and Prevotellaceae were depleted in the diseased subjects, whereas Lachnospiraceae, Ruminococcaceae, Rikenellaceae, Porphyromonadaceae and Bacteroidaceae were more abundant (Costello et al., 2015). Studies on other SpA types have identified *Faecalibacterium prausnitzii* to be decreased in diseased groups (Stoll et al., 2014). The identified taxa may serve as markers because species from these taxa are known to degrade mucin and produce SCFAs. These markers can be potentially associated with AS susceptibility as these taxa may influence barrier integrity and inflammation mediated through Th17 and IL-23 immune signalling (Yang et al., 2016b). A probiotic study on AS patients have been carried out with *Streptococcus salivarius*, *Bifidobacterium lactis*, and *Lactobacillus acidophilus* acting as a consortium of bacteria, which reported no significant clinical improvement (Highton et al., 2012). However, *Bacteroides fragilis* has been reported to benefit AS patients (Stebbings et al., 2009).

3.2.3 Psoriatic arthritis

Psoriatic arthritis (PSA) is a chronic inflammatory arthritic condition that affects the synovial tissue, skin, and the skeletal system along with other clinical conditions like enthesitis, dactylitis and other comorbidities (Veale and Fearon, 2018). PSA is considered a part of the SpA group of conditions and is seronegative (Kruithof et al., 2005;Ritchlin et al., 2017). Onset of PSA in different individuals have varied tissue origin, heterogeneous presentations, different disease routes and outcomes. Studies on PSA have primarily focussed on skin, joints, and entheses (Veale and Fearon, 2018).

PSA is characterised by influx of immune cells and inflammation due to production of pro-inflammatory cytokines in the synovium and damage to bone and other tissues (Veale et al., 1993;Kruithof et al., 2005). The osteoclast cells are also activated which leads to resorption of bone resulting in bone erosion (Sucur et al., 2017). Inflammation at the entheses is the most common characteristic of PSA (McGonagle et al., 1998). The inflammation observed in the DIP joints through MRI and clinical results shows a marked difference in the pattern of inflammation and damage compared to RA. PSA is underdiagnosed in patients with psoriasis due to lack of effective screening methods and lack of proper identification of PSA symptoms. However, early diagnosis and treatment can lead to remission of PSA symptoms (Liu et al., 2014).

The ClASsification criteria for Psoriatic ARthritis (CASPAR) were developed by Taylor *et al.* in 2006 and are the most commonly used criteria, with a sensitivity and specificity of 0.91 and 0.98 respectively (Taylor et al., 2006). This scheme includes the history of psoriasis from the patient or closely related relatives, psoriatic nail dystrophy, swelling of digits and evidence of seronegative and radiographic changes of hands and feet. Various clinical features like axial skeleton disorder, inflammation of the peripheral joints, entheses and dactylitis are indicative of PSA onset and development (Gladman et al., 1987). These conditions can be observed individually or in combination with other factors. Spondylitis, inflammation in the neck and sacroiliitis are the primary features and most commonly identified manifestation in patients with PSA (Helliwell, 2004;Feld et al., 2018). Diagnosis is carried out by radiography which is used to detect bone turnover, asymmetrical distribution and DIP involvement with deformity (Mease and Goffe, 2005). Ultrasonography is an important and widely used tool to evaluate PSA status and efficacy of treatment (Sankowski et al., 2013). MRI is used to identify and monitor

synovial inflammation in PSA and is more sensitive in detecting small damages to bone (Schwenzer et al., 2010).

Like other types of arthritis discussed earlier, various genetic and environmental factors have been identified with the onset and development of PSA (Stuart et al., 2015). Heritability studies have identified strong genetic components with PSA (Myers et al., 2005). Studies with HLA alleles have identified specific HLA susceptibility with subtypes of PSA (Haroon et al., 2016). Excluding the HLA loci, IL-12B, IL-13, IL-23R and TNFAIP3 are also implicated in the pathogenesis of PSA amongst which IL-23R has shown the strongest association (Huffmeier et al., 2010;Eder et al., 2011;Sherlock et al., 2012;O'Rielly and Rahman, 2014;Stuart et al., 2015). High levels of antibodies against Streptococcal infection have also been observed in PSA patients along with evidence of *Streptococcus* genetic material in blood and synovial fluid (Muto et al., 1996). In HIV-positive patients, a depletion of CD4+ T-cells is observed which triggered exacerbation of PSA (Mijiyawa et al., 2000). Amongst various environmental factors, smoking, infection, trauma, and stress have been associated with PSA onset (Mease and Goffe, 2005;Duffin et al., 2009). Smoking is shown to be a risk factor for onset of PSA in people without psoriasis and is considered to be protective in cases of psoriasis which has been associated with polymorphisms in IL-13 (Duffin et al., 2009).

Inflammation at the synovium by various immune cells drives the pathogenesis in PSA. Formation of pannus (abnormal fibrovascular tissue layer) including increased fibroblast-like synoviocytes and activation leads to destruction of the bone and cartilage in the joints (Veale and Fearon, 2018). The lining of the synovium develops increased blood vessels in PSA, whereas RA is characterised by a thick, avascular

lining in the synovium (Ceponis et al., 1998). The synovium of PSA patients also shows an increased expression of growth factors in early stages in contrast to RA that suggests distinct vascular morphology regulation (Fraser et al., 2001;Fearon et al., 2003). The environment of the synovium allows immune cell invasion, proliferation and production of pro-inflammatory cytokines resulting in destruction of the joints and increased risk of associated comorbidities (Ceponis et al., 1998;Biniecka et al., 2010;Biniecka et al., 2016;Hotamisligil, 2017). Secretion of IL-22 induces osteoclastogenesis and activates fibroblast-like synoviocytes, which can increase invasion, contributes to osteoclastogenesis, produce pro-inflammatory cytokines, and produce matrix-degrading enzymes (Mitra et al., 2012). Remission of PSA is achieved by depletion or absence of T-cells by targeting CD4+ and CD8+ through therapeutics. Studies have shown that PSA pathogenesis is marked by increased CD8+ cells compared to CD4+ cells (Veale et al., 2005). The DCs can activate T-cells and in PSA, immature dendritic cells can induce T helper cells to produce TNFα, interferon γ and IL-12 (Candia et al., 2007) and also increase the population of CD8+ T-cells in the synovium (Ramos et al., 2016). Macrophages are also responsible for pro-inflammation and damage to the bone as they secrete large amounts of pro-inflammatory cytokines and matrix metalloproteases as well as induce bone resorption and antigen presentation of T-cells and B-cells (Kurowska-Stolarska and Alivernini, 2017). All these inflammatory reactions result in degradation of cartilage and bone, which have been noted to be architecturally distinct from RA.

Treatment of PSA is aimed at alleviating suffering, preserving the joints, and achieving remission status. Use of glucocorticoids is effective in mild forms of the disease, however long-term use is not advised due to lack of information regarding long-term efficacy, risk of adverse side effects and relapse after discontinuing.

Similarly, NSAIDs are also initially prescribed to patients for short-term treatment of symptoms (Nash and Clegg, 2005). DMARDs are used for treatment of moderate to severe forms of PSA, sometimes in combination with NSAIDs. Conventional DMARDs reduce levels of pro-inflammatory cytokines and inhibit T-cell activity (Cuchacovich et al., 2012). The use of biologics have been suggested as the first line of therapy despite being expensive. Antibodies have been designed to target specific ILs (IL-17A, IL-12 and IL-23), TNF or their respective receptors (Carneiro et al., 2013;Coates et al., 2013;Palfreeman et al., 2013;Cantini et al., 2016). Several anti-TNF drugs are effective in managing PSA but their association with increased risk of infection and tumour formation are highlighted as an adverse effect (Glintborg et al., 2011;Ash et al., 2012;Ali et al., 2013;Minozzi et al., 2016).

Several studies have investigated the link between microbiota, inflammation, and various arthritic conditions. However, extensive investigation with PSA is limited. Scher *et al.* reported microbiota-associated changes in PSA patients for the first time (Scher et al., 2015). Species belonging to *Akkermansia*, *Ruminococcus*, and *Pseudobutyribrio* were less abundant in abundance amongst PSA patients (Scher et al., 2015;Codoner et al., 2018). Taxa identified with PSA patients are involved in mucin metabolism and SCFA production, which are necessary for immune homeostasis. This study observed increased levels of bacterial specific IgA antibodies and decreased concentration of TNF super family 11 in the faecal supernatants. The levels of medium chain fatty acids (hexanoate and heptanoate) were also reported to be lower in diseased groups. These alterations may be associated with microbial immune modulation. The group noted that the microbiota alterations resembled those of IBD patients despite the unique changes in immune proteins. A recent study by Shapiro *et al.* investigated a cohort of psoriatic patients that included both PSA and

psoriasis patients and observed a significant difference in the global gut microbiota profile compared to healthy controls but no difference in α-diversity (Shapiro et al., 2019). Shapiro *et al.* identified 8 genera with increased abundance in the diseased group: *Ruminococcus, Coprococcus, Blautia, Dorea, Collinsella, Actinomyces, Christensenella and Faecalibacterium*. At species level, *Ruminoccocus gnavus, Dorea formicigenerans* and *Collinsella aerofaciens* were more abundant in the diseased group. Key functionality associated with butyrate kinase and phosphate butyryltransferase were depleted while LPS biosynthesis pathways were enriched in the disease subjects (Shapiro et al., 2019). It is interesting to note that the time-period since diagnosis or type of treatment had no association with the gut microbiota profile.

The study by Shapiro *et al.* and Scher *et al.* shows certain contrasting results. The latter observed a decrease in α-diversity amongst the PSA compared to healthy controls, which was not observed in the former study. Species of *Ruminococcus* are important taxa with immune-modulatory properties. Both of these studies report contrasting observations in the abundance of this genus. This may be partly due to unclear distinction between psoriasis and psoriatic arthritis samples in the work by Shapiro *et al.* It may also be due to presence of different species of *Ruminococcus* in the samples that contributed to the cumulative genus level abundance. *Ruminococcus bromii* is commonly observed to be enriched in healthy whereas *Ruminococcus gnavus* is increased in diseased individuals with gut inflammation (Kang et al., 2010;Henke et al., 2019).

3.3 Inflammatory bowel disease

Inflammatory bowel disease (IBD) is a group of chronic inflammatory disorders of the GIT. It is a complex multi-factorial condition characterised by chronic inflammation in the intestine. The two principal types of IBD are Crohn's disease (CD) and ulcerative colitis (UC). CD is known to affect any part of the GIT i.e. from mouth to anus, whereas UC results in inflammation and ulceration in the large intestine (colon) and the rectum (Ordas et al., 2012). Despite being different diseases, CD and UC share certain similar symptoms like abdominal pain, diarrhoea, loss of appetite, weight loss, rectal bleeding, and feverishness. Aetiology of IBD is still unknown; however interaction of various genetic, dietary, and other environmental factors with the immune system is associated with the onset and development of IBD (Xavier and Podolsky, 2007;Wilkins et al., 2011). IBD is an autoimmune disorder but no specific blood markers have been identified that distinguishes CD and UC (Bennike et al., 2014). IBD goes through both active and remission phase and treatment is aimed at achieving remission without relapsing to active form.

Primary diagnostic criteria for IBD are visualisation of the intestine using colonoscopy and by obtaining biopsy samples. Measuring the levels of faecal calprotein is also carried out for intestinal inflammation due to its sensitivity but this technique is not unique to IBD (Henderson et al., 2014). To visualise the features of small intestinal walls, magnetic resonance (MR) enterography is also carried out (Yoon et al., 2017). Currently the Montreal classification system is the most commonly used measure for classifying IBD into either CD or UC (Satsangi et al., 2006). As there is no known marker to distinguish disease type, classification into CD or UC is done by noting the location and nature of the inflammation. Inflammation of

the mucosal layer is characteristic of UC whereas CD is known to affect the whole thickness of the intestinal wall.

Onset and development of IBD is due to complex interaction between genetics, epigenetics, environment, immune system, microbiota and diet (Zhang and Li, 2014). Increased incidences of IBD has been observed amongst relatives and 5-23% of patients had first degree relatives with IBD (Ek et al., 2014). Studies have shown 20-25% of heritability of the risk loci. However, the number of susceptible loci is increasing with ongoing research (Loddo and Romano, 2015). The IBD-associated risk variants were reported in patients with European ancestry (Inoue et al., 2002;Leong et al., 2003). The majority of identified loci are shared by UC and CD, whereas a small number were unique to either disease (Jostins et al., 2012). The first risk allele to be identified was NOD2 (Hugot et al., 2001;Ogura et al., 2001). This gene is needed in the NFκB pathway, a pro-inflammatory pathway that is necessary for proper bacterial recognition (Inohara et al., 2003). A variant of NOD2 associated with CD has been reported to supress anti-inflammatory cytokine IL-10 (Noguchi et al., 2009). Polymorphic variants IL genes; IL-23R, IL-12B, IL-10 and IL-1R2 have also been reported along with STAT3 and JAK2 of which the IL-23 pathway is well established in the pathogenesis of IBD (Tremelling et al., 2007;Anderson et al., 2011).

Amongst the various environmental factors associated with IBD, risk associated with smoking has been the most replicated in IBD studies. Other factors include socio-economic status, stress, diet, and medications. Interestingly, heavy smoking is a risk factor for CD development but confers a protective effect against the development of UC (Lakatos et al., 2007). Vitamin D has immune modulatory properties and low levels of vitamin D were observed in patients diagnosed with IBD

(Leslie et al., 2008). Both high and prolonged dosage of NSAIDs have also been reported to increase the risk of IBD (Ananthakrishnan et al., 2012). Diet is considered to play an important role in the management of IBD (Rajendran and Kumar, 2010;Hou et al., 2011). Increased consumption of fat and meat has been linked with increased risk of IBD, whereas increased fruit and vegetable consumption was related to a decreased risk.

Although the pathogenesis is unknown, various studies have highlighted the role of pro-inflammatory cytokines and dysfunction of the immune system in IBD by investigating the T cell associated mucosal immunity. It is hypothesised that Th1 response is associated with CD whereas Th2 response has been associated with UC (Breese et al., 1993;Fuss et al., 2004;Heller et al., 2005). IBD studies have been carried out to understand immune response in relation to barrier integrity, autophagy, UPR and microbial detection (Wehkamp et al., 2003;Rioux et al., 2007;Kaser and Blumberg, 2011;Salim and Soderholm, 2011). The expression of TLRs and NOD are significantly altered in IBD patients. These receptors play a key role in antigen recognition, immune tolerance, and autophagy. Malfunctioning of TLR2 and NOD2 activate and increase Th1 response (Watanabe et al., 2004;Abreu et al., 2005). IBD has also been associated with UPR-induced stress response from the ER, triggering apoptosis (Kaser and Blumberg, 2011). Various IL families of cytokines produced by the immune cells have been linked with CD and UC or both. These cytokines play a central role in accumulation of other immune cells, triggering response against microbes and a cascade of other inflammatory pathways. Th1 cells produce increased levels of IL-2 and IL-12 that induce IFN-γ in CD (Breese et al., 1993;Peluso et al., 2006). IL-23 plays a key role in Th17 activation and has been observed in both UC and CD (Kobayashi et al., 2008). IL-1β and IL-8 are expressed in UC whereas CD

showed a reduced expression level due to traumatic exposure (Marks et al., 2006). Th2 responses in UC produces elevated levels of IL-4, IL-5 and IL-13 (Nemeth et al., 2017;Imam et al., 2018). *In-vitro* studies have shown that IL-13 may have anti-inflammatory effects and have increased levels of IL-6 when IL-13 is not expressed in UC biopsy samples (Bernardo et al., 2012). Despite the evidence, there is debate regarding the Th1/Th2 drivers (Di Sabatino et al., 2012;Globig et al., 2014;Li et al., 2016). Th17 is highly expressed in both UC and CD, producing IL-17A, IL-17F, IL-21 and IL-22 cytokines through IL-6, TFGβ and IL-23 signalling (Zhou et al., 2007;Guglani and Khader, 2010;Zhang, 2018). However, its role in pathogenesis of IBD is being extensively investigated (Ueno et al., 2018).

IBD is treated using immunosuppressant drugs, dietary alterations and lifestyle changes (Carter et al., 2004). However, surgery is the most common option for management and is carried out least 70-75% of CD patients and over 25-30% of UC patients according to World Gastroenterology Organisation Global Guidelines on IBD. Surgery is required if the patient does not respond to medications. Nutritional therapy is implemented in mild to moderate cases of IBD to rest the bowel and reduce bowel movement. This is achieved by cutting down on fibre, dairy products, digestible carbohydrate, and by adhering to an anti-inflammatory diet resulting in less volume of food passing through the GIT but can also result in nutritional deficiencies (Lee et al., 2015). Nevertheless, nutritional modulation achieves between 50% to 90% remission rates in CD amongst children (Afzal et al., 2005). Remission is also achieved using immunomodulatory drugs like azathioprine, methotrexate, biologics and TNF inhibitors (Louis et al., 2014;Herfarth et al., 2016;Adegbola et al., 2018;Rawla et al., 2018). These are often used to treat severe forms of the disease. Corticosteroids are also prescribed for immune-suppressive effects and antibiotics are used to reduce

microbial load in combination with other treatments but has been linked with high risk of CDI and relapse of the disease (Prantera and Scribano, 2009;Waljee et al., 2016).

Because IBD is an inflammatory condition associated with the GI tract, the role of gut microbiota in its pathogenesis is considered important. Research linking inflammation, IBD and gut microbiota is carried out widely across the world. It has been suggested that the microbiota profiles from IBD subjects may have diagnostic potential. The role of microbiota in IBD has been recognised as early as in 1961 by Weinstein *et al.* and Vince *et al.* in 1972 (Weinstein, 1961;Vince et al., 1972). Two large international cohorts MetaHIT and iHMP have been setup to investigate the role of microbiota in diseases, with one of the key diseases being IBD (Qin et al., 2010;Integrative, 2014). iHMP consists of longitudinal sampling from patients over a one-year period. It contains extensive sampling and meta-data from the subjects accounting for biopsies, metabolites, proteomics, transcriptomics, virome, genetics, serological and dietary information.

Germ-free and antibiotic-treated animal models show absence or attenuation of IBD highlighting the important role of gut microbiota in IBD pathogenesis. Development of IBD conditions has been observed when microbiota from IBD patients were introduced into germ-free mice models. It was observed that germ-free wild type mice colonised with donor microbiota did not develop IBD or inflammation, which may be an indicator of genetic susceptibility (Schaubeck et al., 2016). Overall, various taxa have been identified to be enriched in IBD animal models. These include *Bacteroides, Porphyromonas, Enterobacteriaceae, Akkermansia muciniphila* and *Clostridium ramosum* (Gkouskou et al., 2014). Certain species belonging to these taxa are pathobionts, mucin metabolisers and can have pro-inflammatory effects.

A lower taxonomic and functional diversity is observed in the microbiome of IBD patients compared to healthy subjects. A reduction in the abundance of Firmicutes and Bacteriodetes and an increase in Proteobacteria (mainly species of Enterobacteriaceae and Pasteurellacaea) is seen in IBD patients (Ott et al., 2004;Frank et al., 2007;Kang et al., 2010). However, species of *Enterococcus, Clostridium* and *Listeria* were also observed to be more abundant in IBD (Kang et al., 2010). A loss of commensal bacteria primarily known for SCFA production is also observed (Machiels et al., 2014). Prominent among them are the butyrate-producing obligate anaerobes *Faecalibacterium prausnitzii* and *Roseburia hominis*. These taxa have been linked with reduction in disease activity and gut inflammation and their depletion is linked with relapse of IBD. Other known SCFA producers that are depleted in IBD are *Phascolarctobacterium* and *Paraprevotella* (Watanabe et al., 2012). Leuconostocaceae and *Odoribacter* are also reduced in IBD patients (Morgan et al., 2012). *Lactobacillus* and *Bifidobacterium* are reported to confer a protective effect by reducing disease activity with no observed side effects (Kato et al., 2004;Sood et al., 2009). Biopsy studies have identified Enterobacteriaceae, *Bacteroidales*, *Clostridiales*, Pasteurellaceae, Veillonellaceae, Neisseriaceae and Fusobacteriaceae to be enriched in disease samples while *Bacteroides, Faecalibacterium, Roseburia, Blautia, Ruminococcus* and *Coprococcus* were depleted (Gevers et al., 2014). Bacteria coated with IgA modulate Th17-dependent inflammation and higher concentrations of IgA-coated bacteria induced severe colitis (Palm et al., 2014;Viladomiu et al., 2017). Large-scale comprehensive longitudinal analysis of microbial gene expression and other "omics" data have not been carried out until recently.

A recent publication by Schirmer *et al.* investigating the faecal metatranscriptomics profiles of IBD patients has shown that the both microbial

functional potential and the observed gene expression correlated well with the disease activity across timepoints and samples (Schirmer et al., 2018). Taxa such as *Faecalibacterium prusnitzii* showed shifts in gene expression but not in their taxonomic abundance suggesting that taxonomic abundance are not predictive of their activity. Pathways associated with various sugar metabolism (rhamnose, glucuronate, galacturonate) were associated with difference in gene expression from *Faecalibacterium prusnitzii* but did not correlated with the taxonomic abundance across samples. This shows that the gene expression and functionality of taxa can play a greater role than their taxonomic abundance. Different patient-specific *Bacteroides* species were also reported which may indicate certain similar functionality by different species. However, this does not account for unique genes and their expression found in different species of the same genus. Significant IBD-specific transcriptional alterations were observed for *Alistipes putredinis* and *Bacteroides vulgatus*, *Ruminococcus gnavus* and *Roseburia intestinalis*, showing taxonomic variability between patients and over time. *Ruminococcus gnavus* also exhibited variability in expression suggesting a greater role of this taxa in disease-specific changes in IBD. Functions associated with oxidative stress have consistently been reported as increased in IBD while carbohydrate metabolism and amino acid biosynthesis were decreased. Another recent publication by Lloyd-Price *et al.* highlighted IBD-associated alterations related to microbial, serological, biochemical and host factors (Lloyd-Price et al., 2019). Consistent with other research, a decrease in SCFAs and other known taxa were also observed. The study identified a decrease in stability of gut microbiota composition characterised by frequent and extreme shifts in the microbiota throughout the duration of the disease. Network analysis identified various interactions between

taxonomy, functionality and metabolic profiles with host transcription and serological features.

3.4 Multiple Sclerosis

Multiple sclerosis (MS) is an immune-associated neurodegenerative condition of the CNS. It forms demyelinating lesions in the white matter, grey matter, and the optic nerves. Onset of MS is heterogeneous in nature and includes various complex host and environmental interactions. Various subtypes of MS exist which are considered at an active or inactive stage depending on assessment (Compston and Coles, 2008). The initial onset of the disease is considered a CIS. MS can show reversible relapse of the lesion manifestations also known as RRMS. Once the lesions become more developed, causing permanent neurological deficits, the disease is called SPMS. In a small subset of patients, the disease progression continues from the initial onset without remission and in such cases, it is considered as PPMS (Lublin et al., 2014). The assessment of disease progression and lesion activity is determined using MRI. The diagnosis of MS is based on analysing different regions of the CNS over different periods of time.

Diagnosing MS depends on the clinical manifestation of the disease, as it is heterogeneous in nature with demyelination occurring at different locations. The onset (CIS) is typically characterised by dysregulation of the neural cells, which results in unpredictable loss of functionality, depending on the site of occurrence (optic nerve, spinal cord, cerebrum or cerebellum) in 85% of patients (Miller et al., 2012). EDSS is the most widely used measuring criterion for disability that scores the clinical manifestation from 0 – 10 (0 is a completely normal neurological test and 10 is mortality due to MS) (Kurtzke, 1983). Diagnosis of MS is based on McDonald's 2017

revised criteria (Thompson et al., 2018). In summary, it involves monitoring the disease, disability progression, at least one to two clinical relapses, and evidence for two or more lesions in distinct locations based on neurological examination. Diagnosis also include counting the number of lesions in different areas over time. MRI is also recommended, as it is highly sensitive in detecting defects in the brain and spinal cord (Rovira et al., 2015).

The primary cause of MS is still unknown; however, various genetic and environmental risk factors have been associated with susceptibility to MS. In monozygotic twins, the heritability risk is 35%, whereas dizygotic twins and siblings show 6% and 3% respectively (Compston and Coles, 2002). Amongst the various genes identified as risk factors, polymorphism of HLA class I and II genes are the most prominent (Brynedal et al., 2007;Moutsianas et al., 2015). Polymorphism in T-cell and other immune cell associated genes (IL-2RA, IL-7R and TNF) are also identified as important risk factors for MS (Gregory et al., 2007;International Multiple Sclerosis Genetics et al., 2007;De Jager et al., 2009). It is hypothesised that pathogenesis of MS is also associated with infectious diseases due to involvement of the immune system. EBV infection is the most well established risk factor as epidemiological studies show biomarkers for EBV in 100% of MS patients (Haahr et al., 2004;Olsson et al., 2017). Other factors identified include low vitamin D levels, adolescent obesity, and smoking. However, the exact mechanism of actions is not clear for any of these risk factors (Olsson et al., 2017).

Demyelination which results in lesions is the primary features for all types of MS. This breakdown of the BBB results in influx of active immune cells like macrophages, T-cells and B-cells which escalate the inflammation leading to neuro-

axonal degeneration (Ortiz et al., 2014). Lesions are found throughout the CNS and can affect both white matter and grey matter (Gilmore et al., 2009). The immune cells interact with microglia and astrocytes secreting a variety of pro-inflammatory cytokines that not only induce demyelination causing neural damage but also signal the influx of other immune cells into the area (Li et al., 2018). Insufficient functioning of the T-reg cells results in aberrant T-cell modulation, thus mediating demyelination (Viglietta et al., 2004;Venken et al., 2008;Kitz et al., 2018). Various pro-inflammatory cytokines identified to be involved in MS are IL-17, TNF, IL-12, IL-6, IL-1β and IL-23 (Palle et al., 2017).

MS patients are prescribed DMTs as the first line of therapy as soon as they are diagnosed. Commonly used DMTs are Rituximab, IFN-β or glatiramer (Tintore et al., 2019). In non-responsive patients, transplantation of haematopoietic stem cells is recommended which have proven to be effective (Muraro et al., 2017). More recently, newer biologics have been developed to reduce inflammation rapidly by targeting the T-cells and other immune cells (Buc, 2018). It is interesting to note that the older DMTs have a lower risk profile compared to the newer line of therapy. These newer DMTs are associated with an increased risk of infection (Comi et al., 2017).

Investigating the role of the microbiota in MS pathogenesis in mouse models (EAE) has shown promotion of Th1 and Th17 responses when bacterial adjuvant with myelin protein is introduced (Berer et al., 2011;Mirza and Mao-Draayer, 2017). EAE mice did not develop MS when maintained under germ-free conditions. Under germ-free conditions, animals showed low levels of inflammatory cytokines and increased population of T-reg cells (Lee et al., 2011). The progression and severity of the disease was also reported to be reduced during early treatment with antibiotics. However, in

later stages of MS, there was no effect of the same antibiotics on the mice (Yokote et al., 2008). Th17-mediated development of EAE was observed upon introduction of commensal gut microbiota into the animals (Berer et al., 2011;Lee et al., 2011). Disease susceptibility was also reported when gut microbiota from MS patients were introduced into germ-free mice (Berer et al., 2017). Studies have highlighted that the gut microbiota may be involved in gene expression related to CNS functionality like activating microglia, myelination and BBB permeability (Mirza and Mao-Draayer, 2017). The levels of neurometabolites can be inferred based on certain gut microbial populations according to a study on young pigs (Mudd et al., 2017). Certain lipids and polysaccharides produced by species of Bacteroidetes such as *Bacteroides fragilis* can induce TLR2 tolerance, T-reg cell promotion, suppress CNS inflammation and prevent demyelination in EAE models (Ochoa-Reparaz et al., 2010;Anstadt et al., 2016). Treatment with SCFAs attenuates EAE through T-reg cells and anti-inflammatory cytokine upregulation (Haghikia et al., 2015). Colonisation by *Bifidobacterium animalis, Prevotella histicola* and *Pediococcus acidilactici* as well as species of *Lactococcus* also reduce disease severity (Chu et al., 2018).

Human studies have highlighted gut microbiota taxa depleted in MS that are also associated with decreased SCFAs production, T-reg cell population and microglia maturation (Mirza and Mao-Draayer, 2017). Taxa that were increased in MS patients were associated with mucin metabolism and were known to promote environment-gene immune response. *Bacteriodes*, *Butyricimonas*, *Collinsella*, *Faecalibacterium* and *Prevotella* were reduced while *Bifidobacterium*, *Streptococcus*, *Akkermansia muciniphila* and the archaeal genus *Methanobrevibacter* were enriched in MS patients (Shahi et al., 2017). An increase in the abundance of *Prevotella* and *Sutterella* with a decrease in abundance of *Sarcina* was observed between DMT-treated patients and

untreated patients (Jangi et al., 2016). Probiotic treatment with strains of *Lactobacillus*, *Bifidobacterium* and *Streptococcus* have shown an increase in anti-inflammatory cytokines in human trials (Tankou et al., 2018). A recent study on a Chinese cohort by Zeng *et al.* showed an increased abundance of *Streptococcus* and a decreased abundance of *Prevotella* in MS patients. Th17 cell population were positively associated with *Streptococcus* and negatively associated with *Prevotella*. The study also observed a positive association between SCFAs and T-reg cells (Zeng et al., 2019). Bang *et al.* built different models using different taxonomic levels and different classification methods for predicting various diseases including MS. The group observed that the MS samples consistently had a high false negative rate (32.18-77.01% depending on classification method) and were misclassified as CRC (Bang et al., 2019). The microbiota signature associated with MS may be complex and may have shared certain similarity with other disorders. This would be make disentangling the role of microbiota in MS more complicated.

4. Sequencing and Bioinformatics - a historical perspective

Sequencing is the process of determining the order of nucleotide or amino acids from a given sample of genetic material or protein. It is a continuous linear representation of the sequence, which is also considered as the primary structure and summarises the monomeric units present in a sequence.

The term "bioinformatics" was given by Paulien Hogeweg and Ben Hesper in 1970 where they defined it as the study of information from a biotic system (Hesper and Hogeweg, 1970). The application of computers became essential in Biology when sequences of amino acids became available and the manual comparison of multiple

sequences was found to be impractical. The field of Bioinformatics not only involves the analysis of biological data combining concepts from mathematics, statistics, and engineering, but also requires computer programming for development of tools/methods suitable for biological studies. Bioinformatics expanded rapidly due to advances in sequencing technologies, generation of sequence data and the Human Genome project in the mid-1990s (Venter et al., 2001). Computers became essential for scientific studies as they allowed researchers to compute complex interactions, discern patterns and explore large amounts of biological data.

4.1 First-generation sequencing

As the study of DNA, RNA and proteins continued, methods to sequence these biomolecules developed. In the year 1955, Frederick Sanger reported the complete amino acid sequence of the insulin protein (Ryle et al., 1955). Nucleic acids are a sequence of four nucleotide molecules and also a sugar-phosphate backbone known as adenine (A), cytosine (C), guanine (G) and thymine (T). A sequence of these molecules forms a linear strand of DNA which results in the double helix structure of DNA. In case of RNA, thymine is replaced by uracil (U), which is the demethylated form of thymine. Together, they are also known as nucleobases or bases. Walter Fiers and his colleagues successfully published the earliest form of nucleotide sequencing by sequencing RNA to get the first complete gene of the coat protein in 1972 and the complete genome of bacteriophage MS2 in 1976 (Min Jou et al., 1972). In sequencing, the complex 3-D structure of DNA/RNA is represented by a linear sequence of bases like rows of letters on a page, but without any spaces. In the year 1977, Sanger had developed the methodology commonly known as Sanger sequencing for DNA by

adapting the idea from Ray Wu's work (Sanger et al., 1977). In parallel to Sanger's work, Allan Maxam and Walter Gilbert also published a sequencing method in 1977 known as Maxam and Gilbert sequencing (Maxam and Gilbert, 1977). Despite, being initially more popular than Sanger's method, its use declined due to technical difficulties and improvement in other methods.

Sanger's method remained popular for decades. Sanger's method relies on a chain-termination step. This was achieved by adding ddNTPs of A, T, G, C bases into the reaction pool containing the DNA molecule to be sequenced that acts as the template while bases are added sequentially. Initially the ddNTPs were radioactively labelled however in later version, fluorescent tags were used. Using this method, Sanger and his colleagues successfully sequenced the human mitochondrial DNA of length 16,569 base pairs (bp) in 1981 and the genome of bacteriophage λ of length 48,502 bp in 1982 (Sanger et al., 1982). Subsequently, other groups successfully sequenced the complete DNA of various species of bacteria and other organisms, like yeast. With increased complexity and size of genomes being sequenced, various modifications were implemented into Sanger sequencing to make it more cost effective and faster in order to generate larger sequences. This method was used to sequence the human genome (Venter et al., 2001). The demand for faster, cost-effective methods that could sequence large biological data led to the development of a new generation of techniques, which came to be known as NGS techniques.

4.2 Second-generation sequencing

The second generation of techniques that facilitated sequencing of large biological data on a routine basis was termed as NGS or HTS methods. They had the capability

to sequence multiple individual genomes at the same time. There are four main methods of NGS, namely 1) pyrosequencing, 2) sequencing by synthesis, 3) sequencing by ligation and 4) ion semiconductor sequencing.

In pyrosequencing, at each nucleotide incorporation, light is emitted from a released pyrophosphate (Ronaghi et al., 1996). The type of nucleotide incorporated is identified by the emitted intensity of light. The released pyrophosphate binds to ammonium persulfate releasing an ATP molecule which in turn binds to luciferin. In the presence of the enzyme luciferase, oxyluciferin is produced that generates light of different intensities which is read by a detector. Pyrosequencing can sequence long strands of DNA by adding one base at a time, detecting the light emission, and removing the unincorporated bases, followed by addition of another base and repeating the steps. However, it tends to have a high error rate when strings of homopolymers (~6-mers) are present (Ivady et al., 2018).

Illumina, which is one of the most famous NGS platforms, uses sequencing by synthesis. It makes use of a fluorescent dye and terminated nucleotides as bases (Bentley et al., 2008). All four bases are added at the same time and once a nucleotide is incorporated, the remaining bases are washed off. The fluorescence of the incorporated base is recorded following which the fluorescent molecule and the terminator group are removed. The bases are added again and the whole process is repeated. By repeatedly washing away unincorporated bases after each addition of nucleotide, it is able to overcome the issue with homopolymers encountered by Pyrosequencing. However, it has a high error rate when sequencing long reads because the fluorescent molecules are incompletely washed resulting in a noisier background signal.

The third method is sequencing by ligation, which does not require DNA polymerase to be added with the nucleotides. This technique was developed by Applied Bioscience. In this method, 16-octamer oligonucleotide probes are used that have four different fluorescent dyes at the 5' end of the probes. The octamers contain two probe-specific bases at the 3' end and six degenerate bases. The sequencing starts by hybridisation of the appropriate probe by annealing. The two probe-specific bases guide hybridisation. Unbound oligonucleotides are washed away, followed by recording of the signal given by the fluorescent dye. The dye and the last three bases of the octamer are cleaved, and the next cycle begins. After multiple cycles, the DNA strand is denatured. Another primer is then added to the template strand at one nucleotide upstream position from the previous primer and the steps are repeated. This repetition is done by using five different primers offsetting by 1 each time. This method is limited by its generation of short sequencing reads (Metzker, 2010).

The sequencing process of ion semiconductor sequencing is similar to pyrosequencing; however, instead of pyrophosphate, it detects hydrogen ions released during sequencing to identify the sequences, which is located above a semiconductor transistor (Rusk, 2011). This transistor detects changes in pH due to release of H+ into the solution. This method is much cheaper and faster than pyrosequencing. However, it cannot generate high coverage of large genomes compared to sequencing by synthesis or ligation but is used in clinical setting as it generates sufficient output in a much shorter time period.

4.3 Third-generation sequencing

The third generation of sequencing approaches were first described in 2009 when newer sequencing methods different from second-generation technologies were developed. These methods are still being developed currently. These methods read nucleotide sequences at single molecule level and can generate longer reads compared to second-generation methods (Schadt et al., 2010). These methods can not only generate reads at a much faster rate but can also be used for epigenetic studies (Simpson et al., 2017). They require minimal sample processing and smaller equipment making these technologies portable, which was not possible previously. This allows real time sample collection and data generation, especially in a hospital setting which requires results in a very short time or from extreme places where it is difficult to travel and bring samples back.

Among these third-sequencing methodologies, two of the widely used tools are SMRT from Pacific Biosciences and Oxford Nanopore technologies (Levene et al., 2003;Branton et al., 2008). SMRT utilises ZMW which is a structure that converts electromagnetic waves into light spectra. This method requires a polymerase enzyme, which is attached to the end of a ZMW object. Nucleotide bases are attached with different fluorescent dyes. The ZMW object allows only one nucleotide to be accessed by DNA polymerase and once the complementary base is added to this accessible nucleotide, the fluorescent dye attached to this base is cleaved off. The detector detects the single fluorescence molecule at the time of base incorporation and records it by base calling based on the type of dye.

Alternatively, Oxford nanopore sequencing is dependent on changes in the electrical field inside a nanopore structure when a nucleotide molecule passes through

it. The system contains an electrolytic solution with a constant electric current. When different bases in a sequence pass through the nanopore structure, the magnitude of electric current density changes depending on the type of nucleotide, causing a characteristic electric current density flux (Branton et al., 2008).

5. Processing and analysis of microbiome data

The quality of the sequencing data generated is dependent on various factors like sample processing, amplification, number of samples processed, sequencing depth, base calling and the technology used. Different types of sequencing data are processed in different ways; however, the primary principles remain the same. It involves removing low quality sequences, denoising (correcting errors), removing contamination and identifying taxa and their functional capacity.

5.1 Amplicon sequence data

Certain regions in the genomes of different microbes (bacteria, archaea, or fungi) have been identified to be present across different phyla. These regions are considered as conserved sequences and share homology across species. rRNA and tmRNA are examples of highly conserved sequences. Based on these conserved sequences, microbes can be studied using molecular methods that provides information on the phylogenetic relationship between species. Species with highly similar conserved sequences are assumed to be closely related while distant species would have less homology between them in these sequences. Information on unknown or unannotated taxa can be obtained by reconstructing the phylogenetic relationship based on its sequence homology to known reference species.

In the case of bacteria and archaea, this region is the 16S rRNA gene and the ITS region for fungi (Woese et al., 1990;Baldwin et al., 1995). The 16S rRNA gene codes for the small subunit of the ribosome and is named for the sedimentation coefficient of the product in Svedberg units. The 16S rRNA gene is approximately 1500 bases long. It contains nine hypervariable regions namely V1-V9. The secondary structure of rRNA is composed of helices and loops resulting in both conserved and variable regions (**Figure 3A, obtained with permission**). The length of the different variable and conserved regions and the most commonly used primers and sequencing technologies are also highlighted in **Figure 3B**.

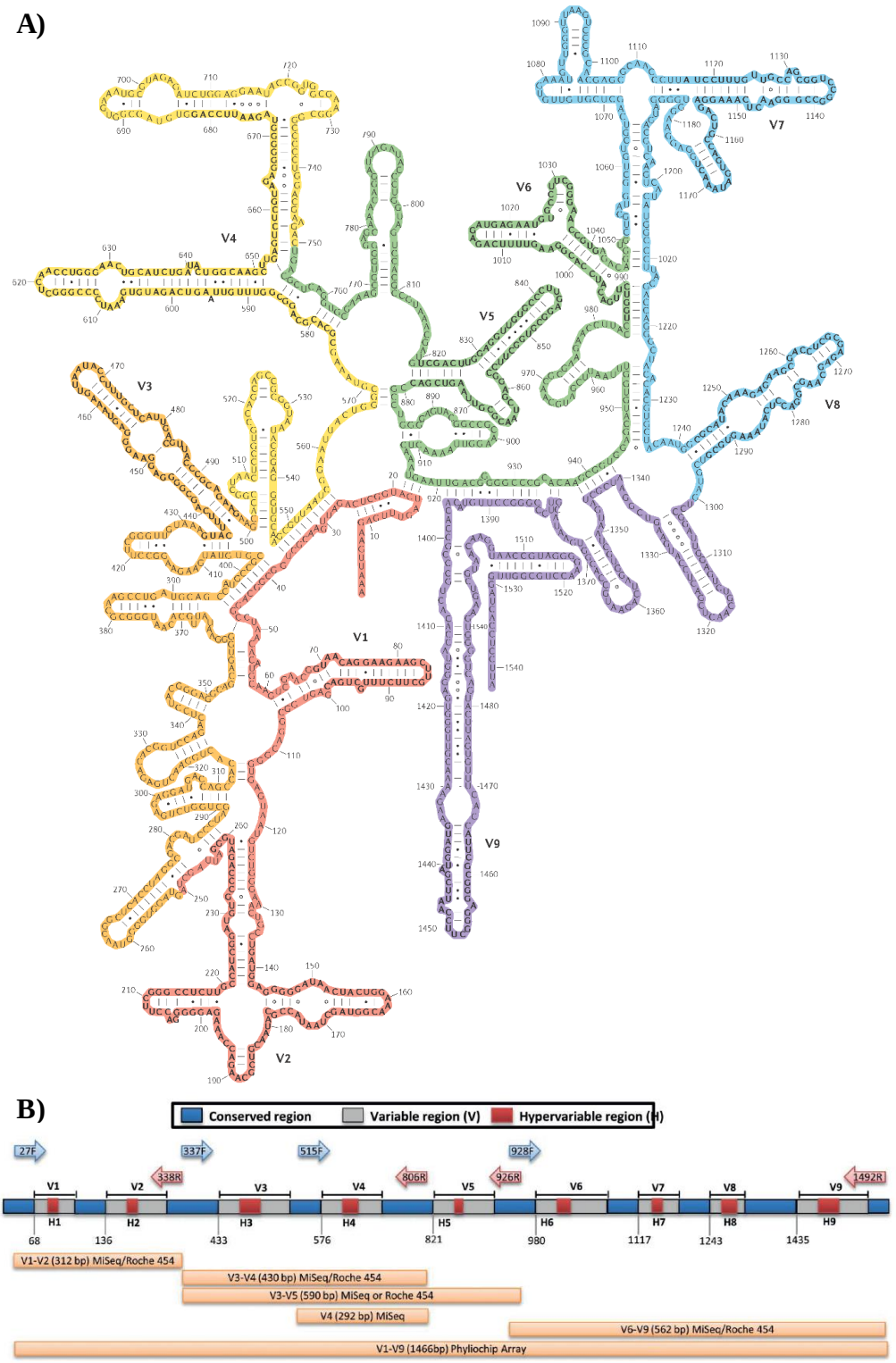


Figure 3A) shows the secondary structure of the 16S rRNA gene, highlighting the hypervariable and conserved regions which form the helices and loops. This figure is obtained from Yarza *et al.* (Yarza et al., 2014) **Figure 3B)** shows the variation in the length of conserved and variable regions along with the most commonly used primers for amplifying the various regions and is obtained from Shahi *et al.* (Shahi et al., 2017).

The 16S rRNA gene is used for phylogenetic reconstruction because it has a slow rate of evolution and is conserved in all species of bacteria and archaea. The use of the 16S rRNA gene for phylogenetic study was pioneered by Carl Woese and George E. Fox in 1977 (Woese and Fox, 1977). The 16S rRNA gene can be used to classify bacteria based on taxonomic specific signature where the 16S sequence is aligned to a reference database. The extent of variability/conservation differs greatly in the hypervariable region as it shows higher conservation at lower taxonomic levels (genus or species) and greater variations at higher taxonomic levels such as phylum or class (Yarza et al., 2014;Yang et al., 2016a). 16S rRNA gene is not only used for identification of bacteria, but it has been used to reclassify bacteria from old phenotypic-based annotation methods and also to classify uncultured bacteria. Although full-length 16S sequences (Sanger sequenced) can be used for greater taxonomic resolution i.e. phylum (low resolution) to species level (high resolution) from a diverse community, it is often not feasible to use due to high sequencing cost. The presence of conserved regions between the hypervariable regions is used to design universal primers that can be used to amplify only a portion of the 16S rRNA gene. By amplifying certain hypervariable regions, it is possible to attain certain levels of taxonomic resolution. Chakravorty *et al.* in 2007 presented the different levels of resolution attainable by the different hypervariable regions and showed that the ability to distinguish different genera or species would depend on the taxa under investigation (Chakravorty et al., 2007).

In microbiota studies, the amplification and sequencing of 16S rRNA gene represents the community-wide population of microbes colonising the sample. As the field of microbiota analysis expanded, numerous studies explored the microbiota community structure, its association with health and disease, and its relationship with

diet and host-associated factors. Large-scale microbiota consortia such as the HMP and the EMP used 16S rRNA gene sequencing and analysis to study microbiota communities (Turnbaugh et al., 2007;Gilbert et al., 2014).

5.1.1 Processing 16S microbiota sequence data

Analysis of microbiota based on 16S rRNA gene sequencing data has its own computational challenges. During the early phases of 16S rRNA-based microbiota studies, Riesenfeld *et al.* in 2004 reported that Genbank had 21,466 and 54,655 16S rRNA gene sequences from cultured and uncultured prokaryotes respectively (Riesenfeld et al., 2004). Based on 16S rRNA analysis only, many candidate phyla were proposed; however, many of these phyla had no cultured representatives. Over the years, various pipelines have been developed for processing and analysing 16S rRNA microbiota data. Some of the most popular and widely used pipelines and tools are QIIME1 and QIIME2, mothur and USEARCH suite of tools (Schloss et al., 2009;Caporaso et al., 2010;Edgar, 2010;Bolyen et al., 2019). These pipelines/tools incorporate various steps required to generate a taxonomic abundance table and various diversity measures.

Sequence data generated from 16S rRNA genes can be several GB in size. Processing the 16S sequencing data to obtain the taxonomic abundance data requires multiple steps that involve removal of adapters and primer sequences, quality filtering of the sequence data, merging forward and reverse paired reads to obtain the full length of the hypervariable region and generation of a taxonomic abundance table. The removal of primers is necessary because primers contain ambiguous characters which allow matching to multiple nucleotides in the templates. Primer inclusion can allow

mismatches to occur thus introducing artificial variations in the hybridised region. Removal of these sequences can be carried out primarily in two ways. Software such as cutadapt (Martin, 2011), split_libraries.py (QIIME) or truncate_reverse_primer.py (QIIME) can be used to trim the sequences by using adapter/primer sequences as references. The software identifies adapter and primer regions allowing a user-defined number of expected nucleotide differences. The advantage of this method is that you can trim the adapter/primer region of interest accounting for the differences in length and nucleotide sequence due to errors. Alternatively, one can also trim by length which removes bases of a specified length, ideally the length of adapters/primers from the beginning of the sequenced reads. Quality filtering of the reads is done by examining the phred scores (a measure that defines the quality and accuracy of bases calling) of the reads and removing those that fall below a certain threshold. A sliding window approach is often used to calculate the average read quality over a number of bases and trimmed if the average quality drops below the threshold. Trimmometric can be used for both primer removal and trimming purposes (Bolger et al., 2014). Quality filtering ensures that low quality reads are removed, thus retaining a filtered set of reads, whose nucleotides can be attributed to high quality base calling.

In paired-end sequencing, forward and reverse 16S reads must be merged to obtain the full-length hypervariable region of interest. Hence, it is recommended that the overlap between the paired reads should be large (e.g. over 10 bases for V3/V4 region) (Schirmer et al., 2015). The length of overlapping regions is dependent on the length of amplicon sequenced and the read length obtained from the sequencing technology used. Large overlapping regions are also particularly helpful as quality filtering normally trims few a bases at the beginning and at the end of reads due to poor quality which can reduce the region of overlap between forward and reverse

reads. Large overlaps also ensure correct length of the merged sequences. The success of overlap is also dependent on the quality of base calling. Merging of paired end-reads can be done using tools like Flash, PANDAseq and PEAR that effectively merge reads over a relatively short period of time (Magoc and Salzberg, 2011;Masella et al., 2012;Zhang et al., 2014). Flash can calculate the length of overlap based on the input reads. The proportion of reads merged can be tuned by adjusting different parameters such as the acceptable range of expected nucleotide errors, the length of sequence overlap and expected length of the merged sequences (including standard deviation in length).

The quality-filtered, merged reads are processed/clustered to determine OTUs i.e. representative sequences which can be assumed to be similar enough to be considered as OTUs (Blaxter et al., 2005;Schmidt et al., 2014). A representative sequence is chosen from among the reads within each OTU by one of several methods (e.g. most common, or centroid). The idea behind clustering is to reduce the effect of errors due to incorrect base calling that can result in nucleotide polymorphisms. This exaggerates the number of unique sequences, resulting in a greater number of false positive and thus a higher measure of community diversity. Two different approaches were developed to address this concern. Closed reference clustering and *de novo* clustering.

In closed reference clustering, the reads obtained from sequence data are mapped to a reference prokaryotic database (Greengenes) (DeSantis et al., 2006;Rideout et al., 2014). Here, each reference sequence in the database takes the place of an OTU and all reads mapping on to that sequence above a user-defined threshold (i.e. 97% sequence identity) are placed in the same taxon group. This

allowed reads to be successfully mapped to a set of known representative taxa and provided a curated set of representative bacterial taxa. However, by using this method, identification of novel or phylogenetically distant taxa is not possible. This may result in discarding a large volume of sequenced reads as "unknown" and is not convenient for analysing samples from sites/environments that do not contain well-studied/representative taxonomic diversity.

To overcome the limitation of a lack of representative sequences, the *de novo* method was developed (Rideout et al., 2014). In the *de novo* method, the reads from the sequencing run are clustered based on a similarity threshold (e.g. 97%) i.e. all sequences, which are similar by at least 97% or more, are clustered together and represented by one representative sequence (considered the centroid). As reads are not aligned to an external database, potentially all the reads in the samples can be used thus ensuring detection of novel or rare taxa. However, given the fact that the reads are not mapped to a reference database, it is possible that representative sequences which are not biological in nature but are formed as a result of amplification error would remain. This may result in generating false OTUs e.g. chimeric sequences or chimeras.

Chimeras are hybrid sequences that arise due to incomplete amplification of the template DNA during a PCR cycle (Haas et al., 2011). Such sequences can get amplified in the next cycle by merging with another sequence which is not its original template. These hybrid sequences thus are a result of two different templates that proceed with further amplification as the PCR cycles continue. Factors that mainly influence generation of chimeric sequences are enzymes used, PCR cycles and level of sequence identity between 16S rRNA sequences. It has been estimated that

chimeras can represent 15% of the total reads in a study or even higher (Gohl et al., 2016). Chimeric sequences pose a greater problem during *de novo* methodology due to lack of a reference database which can also be used to distinguish biological sequences from chimeric ones. To detect and remove chimeras, two different methods have been proposed. The first method utilises a curated chimera-free reference database to which the representative OTUs are mapped using a tool, ChimeraSlayer (Haas et al., 2011). It compares all sequences and searches for potential hybrid sequences that may have formed from multiple sequences to the database and removes those sequences which are considered chimeras. Alternatively, *de novo* methods were also developed to detect chimeras. UCHIME is an algorithm that implements a *de novo*-based chimera detection method by investigating the representative sequences in decreasing order of their abundance (Edgar et al., 2011). The references sequences (parent OTU) must be 2 times abundant than the query sequences (sequence investigated). The query sequences are fragmented and aligned against the various parent OTUs to identify if the query sequences belong to different parent sequences or not. Recently Edger showed that UCHIME has a lower error rate than reference-based methods, which is likely to remove false positive predictions (Edgar et al., 2011). He also showed that no method in principle can distinguish all chimeras accurately from true biological sequences (Edgar, 2016).

A hybrid method of both closed and *de novo* OTU generation known as open-reference OTU clustering can also be implemented (Rideout et al., 2014). In this method, the filtered reads are first processed using a reference-based methodology and reads excluded from this step are subsequently clustered with *de novo* method. This strategy may be preferred over the previous methods as it not only allows maximum utilisation of the reads but also generates a set of reference based OTUs with known

taxonomy. Once the representative OTUs are identified then all the reads from the sequencing data are mapped to these OTUs. The number of reads mapped to each OTU is tabulated which results in an OTU count table.

One of the key concerns with representative OTU sequences is that they may be underrepresenting certain taxa which could be similar by more than 97% percent e.g. species from the *Enterobacteriaceae* family (Jovel et al., 2016). With improvements in sequencing methods and generation of high-quality sequencing data, alternate methods have been developed that do not rely on clustering. Examples of these methods are oligotyping, DADA2 and Deblur (Callahan et al., 2016;Amir et al., 2017). These methods can identify variants in sequences down to a single nucleotide polymorphism. In oligotyping, site-specific nucleotide polymorphisms can be detected which can then be referenced and would allow discrimination of closely related species and strains (Eren et al., 2013;Eren et al., 2015). DADA2 resolves unique sequences, which are also termed as SVs, at high resolution by calculating the transition and transversion rates between the nucleotides and also considering the observed and expected frequency of nucleotides at each position from the given sequence data. This is done in order to distinguish true sequence variants from sequencing errors and thereby correct the errors (Callahan et al., 2016). This method does not rely on reference databases so reference-based chimera detection is not an option, but can be included as a later step in the 16S processing pipeline.

OTUs and SVs generated through clustering approaches (uclust), DADA2 or Deblur lacks taxonomic annotation. Taxonomic annotation would allow classification of the OTUs and SVs at phylum, class, order, family, genus and/or species levels. Various methodologies have been developed over the years for classification of

amplicon sequences. The most popular of these are the Ribosomal Database Project (RDP)-classifier, UTAX, SortMeRNA, UCLUST and nucleotide BLAST (Edger;Wang et al., 2007;Kopylova et al., 2012). These tools allow reliable classification of sequences down to genus level. RDP-classifier, UTAX, and SortMeRNA use k-mers to compare query sequences to a reference database. K-mers are substrings of length k obtained from a representative OTU sequence or SV that are then compared with k-mers from a reference sequence – the more k-mers two sequences share, the higher the confidence/similarity score. BLAST locates exact matches of length k (8 by default) and extends these in both directions (5' and 3') based on a substitution matrix until the score drops below a user-defined (or default) threshold. This heuristic approach is similar to the Smith-Waterman algorithm but is magnitude faster at the expense of accuracy. However, species-level classification of 16S data is difficult due to the very high levels of similarity between species of the same genus. New tools have been developed in recent years that allow classification at species level: SPINGO, BLCA and IDTAXA (Allard et al., 2015;Gao et al., 2017;Murali et al., 2018). SPINGO also utilises a k-mer approach and a curated RDP reference database annotated to species level. IDTAXA also uses a k-mer approach but provides multiple databases from which to choose. BLCA assigns taxonomy based on Bayesian posterior probability and calculates the likelihood of taxonomic assignment. This likelihood is defined first by BLAST alignment of query sequences to reference databases, followed by multiple alignment using muscle. Bootstrapping of the alignment is then performed 100 times to calculate the posterior probability of each hit and to obtain a confidence score for each best hit.

All these taxonomic assignment methods are dependent on various databases. The databases commonly used are RDP database, GreenGenes and SILVA (DeSantis

et al., 2006;Quast et al., 2013;Cole et al., 2014). The RDP database consists of 16S rRNA gene sequences obtained from INSDC and the current version (11.5) has 3,356,809 16S rRNA gene sequences. The GreenGenes database is smaller than the RDP database; however, sequences present in this database are filtered to remove chimeras. It provides representative sequences of clusters of 16S rRNA genes at a different identity threshold - the latest version, 13_5, contains 203,452 representative sequences at 99% identity. The taxonomic annotations of the 16S rRNA gene sequences in the SILVA database are obtained from Bergey's Taxonomic Outlines and the List of Prokaryotic Names with Standing in Nomenclature (Parte, 2018). All assignments are manually curated. The latest update version, 132, contains 617,630 non-redundant 16S rRNA gene representative sequences (also taken from 16S gene clusters at 99% identity).

5.1.2 Downstream data analysis

The OTU/taxonomic data generated are analysed using various techniques some of which are adopted from macro-ecological or RNA-Seq studies. However, prior to applying any methodology, it is important to ensure that the data from all the samples are scaled/normalised to an even level to make them comparable. This is necessary as a sequencing run usually results in an uneven distribution of read depth across samples due to the stochastic nature of the sequencing process. Rarefaction is a commonly used normalisation technique in which the reads from each sample are randomly subsampled (without replacement) to a uniform sequencing depth (Hughes and Hellmann, 2005). However, this results in discarding valid data from samples with greater sequencing depth which has led to criticism about rarefaction being

statistically unacceptable (McMurdie and Holmes, 2014). Aside from rarefaction, other approaches have also been found to be applicable for microbiota data. These methods are CSS, proportional normalisation, UQ scaling, DESeq-VS, regularised log and log transformation on proportional or rarefied data (Bullard et al., 2010;Robinson et al., 2010;Paulson et al., 2013;Love et al., 2014;Weiss et al., 2017;McKnight et al., 2019).

One of the primary and key types of analysis of microbiota data is investigation of diversity. Diversity analysis involves within-sample diversity or α-diversity and between-sample diversity or β-diversity. The α-diversity of a microbial community is a measure of taxonomic diversity and is determined using various indices including species richness, Chao1, Shannon index, Simpson index and phylogenetic diversity (Simpson, 1949;Whittaker, 1972;Faith, 1992;Spellerberg and Fedor, 2003;Chao et al., 2006). Species richness is a measure of the number of unique species-level taxa (often the OTU count) detected. Chao1 is an extrapolation of detected species to estimated total species giving more importance to low-abundant OTUs such as singletons and doubletons in each sample. Shannon and Simpson indices calculate diversity by accounting for the richness and taxon abundance in each sample (species evenness). Shannon gives more weight to richness whereas Simpson gives more importance to species evenness. Phylogenetic diversity calculates the total branch length for all OTUs in each sample from a pre-generated tree built from all representative OTU sequences.

The β-diversity of a microbial community is a measure of dissimilarity/distance between samples calculated from the taxonomic abundances of samples (Whittaker, 1960). Dissimilarity measures range from 0 to 1. A value of zero

denotes identical representation of taxa between two samples and a value of 1 denotes complete dissimilarity of taxonomic profiles between two samples. This can be used to define the microbiota community structure of the samples in a dataset. BC, Jaccard, Weighted and Unweighted UniFrac distance are some of the metrics used to calculate the β-diversity between samples in microbiota studies (Jaccard, 1901;Bray and Curtis, 1957;Lozupone and Knight, 2005). BC distance is measured by 1 – BC similarity. BC similarity is defined as the sum of the abundance of taxa/OTUs shared by two samples divided by the number of taxa/OTUs detected in both samples. Bray-Curtis considers both richness and evenness in its calculations. Similarly, Jaccard distance is calculated by 1 – Jaccard similarity. Jaccard similarity is defined as the number of common taxa/OTUs detected between two samples divided by the total number of taxa/OTUs detected. Jaccard does not consider the abundance of taxa/OTUs in its calculation. UniFrac distance is a more modern metric and was developed for microbiota analysis - it is one of the most commonly used β-diversity metrics. UniFrac distance calculates the shared branch lengths between two samples and expresses this as a fraction of total branch length. Two variations of this method are Weighted UniFrac in which taxonomic abundance is considered and Unweighted UniFrac which only accounts for the presence or absence of taxa. The β-diversity measures are commonly visualised using dimension reduction methods such as PCoA or dbRDA. The effect size of metadata groups (e.g. healthy vs diseased) and other factors and their significance at explaining the variation in the global community profiles can be calculated using permutational-based approaches such as PERMANOVA or randomisation using Monte-Carlo permutation.

Another key analysis of microbiota data is identifying differences in taxonomic composition that are associated with changes in diversity and groups. These

differentially abundant taxa serve as key markers that are associated with different groups/conditions. Identification of the differentially abundant taxa is done using various statistical methods. Due to the non-normal distribution of the microbiota data, non-parametric methods are commonly used to identify significant taxa where raw values are converted into relative ranks. Some of the simplest approaches include Wilcoxon Signed-Rank test, Mann–Whitney U test, Kruskal-Wallis test, and Dunn's test. Given that the detection of differential taxonomic abundance can be inaccurate due to the influence of uneven sequencing depth across samples, it is necessary to normalise the data as discussed before. The datasets are also zero-inflated due to a spare distribution of most taxa; hence filtering of the dataset to remove very low-abundant taxa with a majority of zeros is also carried out before identifying the differentially abundant taxa. Presences of zero can represent the absence of a taxon in a sample or its presence below the limits of detection. They are removed because they break all the assumptions of statistical tests, although this depends on the test. Identification of significant taxa is dependent on the sensitivity of the method implemented. Non-parametric statistics are usually less sensitive than their parametric counterparts. To capture the biological signal accurately, various methods have been designed or adapted from RNA-Seq studies. These methods rely on various normalisation procedures. DESeq2 uses Wald's test and uses a negative binomial distribution for the data, modelling taxa count with fitted mean and feature-specific dispersion parameter (Love et al., 2014). DESeq2 also handles outliers using Cook's diagnostic test and replaces them with the trimmed mean over all groups if at least seven samples are present in a group. MetagenomeSeq is another methodology designed specifically to deal with sparseness in microbiota datasets (Paulson et al., 2013). This methodology uses CSS normalisation, which calculates a scaling factor

by dividing the sum of the read counts up to a particular quartile for a sample and applies a mixed model with zero-inflated Gaussian distribution that can account for varying sequencing depth and under sampling. It also applies Fisher's exact test to presence-absence data and odd's ratios. ALDEx2 is another recently published robust methodology for analysing high throughput sequencing data (Fernandes et al., 2013). It uses Bayesian approaches to account for technical variations and implements a clr transformation.

Due to the compositional nature of microbiota data, the use of different normalisation and abundance testing strategies have been debated. McKnight *et al.* have shown that proportional transformation can be utilised for investigating community structures and agrees better with actual data compared to other transformations (McKnight et al., 2019). Weiss *et al.* have compared different normalisation techniques highlighting the different levels of sensitivity for various methods when analysing differential abundance (Weiss et al., 2017). The group reported that rarefaction showed better segregation of the samples based on microbial community structure. DESeq2 showed high sensitivity with smaller datasets however was affected by large difference in library size.

Amplicon sequencing is extensively used for microbiota studies due to cheaper cost, smaller data size and less computation time. It is also very informative when analysed properly. Despite these advantages, it has certain limitations. Currently, sequencing of the full-length 16S rRNA gene with sufficient coverage to capture all microbial species from a large number of samples is usually not possible from 16S based microbiota data. In terms of amplification and sequencing, most commonly used primers cannot effectively target archaea (Acinas et al., 2005). Presence of multiple

copy number of 16S rRNA gene in a genome, low amplification from microbes from low biomass samples and sequencing bias can affect can affect the accurate estimation of the microbial population (Kembel et al., 2012;Karstens et al., 2019). Variation in genomic copy number of 16S gene will affect the estimation of relative abundance of species. Contamination during sample collection and processing are amplified during PCR and gets sequenced to a greater sequencing depth dominating low microbial biomass. This results in false inflation of diversity and abundances. To control for contamination, negative controls are processed alongside, and any microbial load observed in the negative controls are removed from all other samples from the same run. Multiple sequencing runs can also result in run-to-run cross-contamination, thus effecting accurate estimation of the microbiota population. The taxonomic resolution of the representative sequences is down to genus level only with high confidence in many cases. Hypervariable regions vary in their ability to detect taxa at different taxonomic levels. This resulted in different recommendations (Chakravorty et al., 2007;Mizrahi-Man et al., 2013). These differences are due to various factors including use of different primers, sampling environment and classification approaches. It is also not possible to determine directly the functional potential of the microbial community from 16S data only (Langille et al., 2013).

Methods for indirect functional prediction from 16S rRNA data have been developed in which representative sequences/taxa are mapped to the closest functionally annotated genome and the functional abundance is determined based on the taxonomic abundances. PICRUSt is the earliest method to attempt this (Langille et al., 2013). PICRUSt generates a phylogenetic tree of the reference sequences obtained from Greengenes. From the functionally annotated references in the tree, ancestral gene content is inferred from the ancestral branches. This inferred gene

content is then added to the reference sequences with unknown functions. This step ensures that all reference sequences represented by tips on the phylogenetic tree have either known or predicted gene content. This information is precalculated and is not required to be computed every time. To infer functional potential of the queried data, representative OTU sequences are aligned against the Greengenes database, and the OTU identifiers are matched to the reference tree. Functional abundance is calculated by multiplying the gene content of the matched reference sequences with the OTU abundance from the table. One of the limitations of this approach is that phylogenetically distant sequences may have very distinct functionality from their most closely matched reference genomes, thereby making functional predictions for these sequences inaccurate. To overcome this issue, Tax4Fun and Piphillin are two other tools that were developed (Asshauer et al., 2015;Iwai et al., 2016). Tax4Fun utilises the SILVA database as a reference and sequences are processed into KEGG annotated taxonomic profiles using BLAST against KEGG prokaryotic genomes. It uses Taxy-Pro to obtain reference organism-specific profiles (Klingenberg et al., 2013). Using UProC and PAUDA, it computes functional KOs for the annotated OTU table (Huson and Xie, 2014;Meinicke, 2015). Piphillin is a web-based functional prediction (KOs) tool from the start-up company SecondGenome. Unlike PICRUSt, which requires OTUs to be mapped to Greengenes, or Tax4Fun that requires mapping to SILVA, Piphillin does not require any such processing. The representative sequences and OTU table are uploaded to the webserver. The representative sequences are aligned using global alignment against 16S rRNA gene sequences of known genomes. The OTU table is adjusted to the mapped genomes and subsequently normalised for copy numbers variation, followed by functional profiling using the gene content of the reference genomes. While all these tools boast high predictability,

they are still limited to calculating the functional potential for only those OTUs/representative sequences which can be successfully mapped to the functionally annotated reference genomes.

5.2 Shotgun sequence data

Metagenomic whole genome shotgun (mWGS) sequencing is a method of sequencing the total genetic material of a sample. It is a massively high-throughput sequencing process, where the DNA obtained from the samples is sheared into random fragments of smaller read lengths and sequenced (Claesson et al., 2017). It is designed to capture the fragmented whole genomes of the microbial communities. There is no amplification of a specific marker gene; only sequencing adapters and indexing primers are added. It is an expensive method but can generate hundreds of GB to TB of sequencing data, which is magnitudes larger than 16S data depending on the sequencing technology. Hillmann *et al.* showed that sequencing depth of as low as 0.5 million reads can provide similar information as ultra-deep shotgun sequencing (Hillmann et al., 2018). This allows consideration for reduced cost and generation of lower volumes of sequencing data while still capturing the dominant taxa. With a reduction in cost and advancing computational capacity, mWGS is becoming a preferred method for current and upcoming projects. Using shotgun metagenomics, a community-wide functional profile can also be determined, although this is not necessarily indicative of the full functional capacity of the microbiota as it only identifies genes/pathways present in the metagenomic data and not their expression levels (Franzosa et al., 2014;Schirmer et al., 2018).

5.2.1 Processing mWGS sequence data

Due to the large volume of data generated, processing mWGS sequence data is computationally intensive and takes longer than 16S data. Quality trimming of sequence data can be processed in a similar manner as discussed earlier for 16S data. In mWGS sequence data, host genome contamination is a major concern. As mWGS data are processed from the total DNA extracted, and the protocol uses no amplification steps, all DNA fragments including host genomic content are indexed and sequenced. To remove host contamination, reads from the samples are mapped to the host reference genome and all the mapped reads are considered as contamination and removed. This is particularly important when processing samples with low biomass, which includes samples obtained from the host like biopsy samples or skin samples (Eisenhofer et al., 2019). These samples must be processed carefully to avoid contaminated reads during downstream processing which could result in presence of host genes and pathways.

Primarily there are two approaches to analyse mWGS data, assembly-based and reference-based. In assembly-based methods, the whole genomes of the organisms in a community are potentially reconstructed by assembling the sequenced reads into longer continuous sequences known as contigs which are further assembled into scaffolds. The subsequent joining of the reads into continuous longer reads is done using the de Bruijn graph algorithm. This approach is also considered to be a *de-novo* assembly method as no reference genomes are involved. Tools like IDBA (Peng et al., 2010) and Velvet (Zerbino and Birney, 2008) were initially developed to assemble single genomes; however, different variations of these tools (Meta-IDBA, IDBA-UD, MetaVelvet-SL) were developed for metagenomic sequence data which contains

many genomes from a diverse array of microorganisms (Peng et al., 2011;2012;Afiahayati et al., 2015). Other commonly used tools are SOAPdenovo2 and SPAdes (Bankevich et al., 2012;Luo et al., 2012). IDBA-UD applies a depth threshold to eliminate contigs with low depth. MetaVelvet-SL implements a Support Vector machine learning algorithm to identify chimeric nodes (k-mer sequences shared between genomes of similar species). Once contigs and scaffolds have been built, these sequences can be annotated taxonomically and functionally. Binning is a strategy that categorises sequences into groups that are likely to belong to the same genome and is carried out based on nucleotide compositions or coverage. Binning accuracy is partly affected by the length of sequences and reads. CONCOCT, MetaBAT, Maxbin and Maxbin2.0 are amongst the commonly used binning approaches and these tools utilise nucleotide frequency and coverage to obtain different bins where each bin contains sequences from potentially similar genomes (Alneberg et al., 2014;Wu et al., 2014;Kang et al., 2015;Wu et al., 2016). Given that scaffolds and contigs are representative of a full length or partial genome, the functional potential of these sequences can be determined by predicting the genes. Genes can be predicted by a variety of tools like MetaGeneMark, PRODIGAL and GLIMMER (Salzberg et al., 1998;Zhu et al., 2010;Hyatt et al., 2012). The reliability of predicted genes is dependent on the quality of the assembled sequences. Low coverage or a large number of smaller contigs results in incomplete genes that are truncated at either their 5' or 3' ends (or both) by contig boundaries. Predicted genes can be annotated using methods like BLAST against functional databases such as KEGG, UniRef, COG, TIGRFAMs or InterPro (Kanehisa and Goto, 2000;Tatusov et al., 2000;Haft et al., 2001;Suzek et al., 2015;Mitchell et al., 2019). Using assembly-based methods, it is possible to assemble genomes of unknown or novel taxa.

However, as mWGS sequencing data is represented by various taxa and can have various closely related strains, building contigs and scaffolds may collapse such strains into a single representative genome thus preventing strain-level resolution (Breitwieser et al., 2017).

To overcome the limitations of an assembly-based approach, reference-based methods were developed. The reads are mapped to a database directly which would contain taxonomic marker sequences or functionally annotated genes. Various tools have been developed using different approaches for fast and accurate classification using composition-based or alignment-based approaches. Amongst composition-based tools, Kraken and K-SLAM compare the k-mers of queried reads against a reference database (Wood and Salzberg, 2014;Ainsworth et al., 2017). The reference database consists of prokaryotic genomes from the NCBI RefReq database from which the LCA taxa is determined for the k-mers of reads mapped. K-SLAM validates the alignment using local alignment and pseudo assemblies of genomes from similar species. K-SLAM reports both taxonomic and gene information. Kaiju implements the BWT algorithm on 6-frame translated sequences against the NCBI RefSeq database of microbial and viral protein sequences (Li and Durbin, 2009;Menzel et al., 2016). Amongst the alignment-based tools, MetaPhlAn and MetaPhlAn2 map the reads against a clade-specific marker database to obtain bacterial and archaeal taxonomic information (Segata et al., 2012;Truong et al., 2015). MetaPhlAn initially implemented BLASTn as a default but it can also use other tools like MBLASTX (Davis et al., 2015). MetaPhlAn2 is the enhanced version consisting of a larger reference database expanding into viral and eukaryotic markers and also allows strain-level analysis also using StrainPhlAn (Truong et al., 2017). MetaPhlAn2 uses Bowtie2 to reduce its computation time (Langmead and Salzberg, 2012). As reads are mapped

only to known clade specific genes, MetaPhlAn2 can introduce false negatives when generating taxonomic profiles as only known annotated reference taxa are identified. To obtain functional information from reads, tools like DIAMOND can be implemented which aligns short reads to protein databases (Buchfink et al., 2015). The sensitivity of DIAMOND is similar to BLASTX (up to 94%) and it is a magnitude faster than BLASTX.

Due to the availability of a variety of tools and databases, various pipelines have been developed to automate the process of obtaining taxonomic and functional profiles. For assembly-based approaches, MOCAT2 and MG-RAST pipelines have been developed (Meyer et al., 2008;Kultima et al., 2016). These pipelines implement a variety of tools, some of which were discussed earlier, in an automated manner to generate taxonomic and functional profiles. In summary, MOCAT2 uses FastX, SolexaQA and an internal script for quality filtering and trimming (Hannon, 2010). Assembly of quality-filtered reads is carried out using SOAPdenovo. Prediction of genes is carried out by MetaGeneMark and Prodigal. Taxonomic and functional annotation is carried out using SOAPaligner and DIAMOND, respectively (Gu et al., 2013). Amongst reference-based pipelines, HUMAnN and its next-generation HUMAnN2 pipeline has been developed (Abubucker et al., 2012;Franzosa et al., 2018). In HUMAnN2, a taxonomic profile is obtained using MetaPhlAn2. Subsequently a functional profile is obtained using a nucleotide and a 6-frame translated search. It initially maps the reads to a pangenome database of functionally annotated species using Bowtie2. Unmapped reads are processed using DIAMOND as default against the UniRef database to obtain protein profiles from unknown taxa. MetaCyc Pathway profiles are then generated from protein profiles using Minpath (Ye

and Doak, 2009). KEGG orthologs, modules and pathways can also be obtained from the UniRef profiles by using the legacy database of HUMAnN1.

Downstream analysis of the taxonomic and functional profiles can be carried out in the similar manner as discussed earlier, including calculation of α- and β-diversity, and the use of statistical tests/methodology to detect significantly different compositions of taxa and functions.

6. Conclusion

In summary, there is an increasing number of studies on both animals and humans that highlights the relationship between the microbiota and its host. Limitations in culture-based methods underlines the importance of culture-independent and HTS approaches using computational tools. Secretion of antibodies (IgA) and AMPs by immune cells into the mucosa of the gut, SCFA-associated regulation of T-reg cells, production of bacterial polysaccharides which triggers T-helper cells, availability of nutrients and the integrity of the mucosal layer can influence the microbiota and vice-versa. These prominent phenomena suggest a role for the gut microbiota in immune homeostasis and inflammation as highlighted by various germ-free animal models. It is promising to explore these relationships in different immune-mediated disorders. It is also necessary to identify the effect on the microbiota population and their functionality associated with changes in host metabolism, medication, diet, and other factors. However, as various covariates are associated with microbiota, it is necessary to account for these confounders and adjust for their influences. This can be achieved by carrying out both observational and longitudinal studies. Studying immune-mediated disorders will uncover shared and unique microbiota signatures associated with these disorders. Uncovering host-microbe/microbe-microbe interactions and their effects in disease and health will also facilitate development of effective *in-vitro* and *in-vivo* experiments for validation. This is essential for generating a knowledge base necessary for potential personalised diagnostic and therapeutic implementation.

7. Aims and objectives

This chapter summaries the evidences highlighting the impact of the gut microbiota on human physiology. It establishes the gut microbiota as a modifiable component with major implications on the health of human. Despite the progress in microbiota research associated with immune homeostasis, there a tremendous knowledge gap regarding exploiting the malleable nature of microbiota for potential diagnostic/therapeutic purposes in various immune-mediated diseases. For this, it is necessary to understand and establish the microbiota-immune interaction in the aetiology of various immune disorders. The research carried out in this thesis is focused on exploring and establishing the microbiota profiles associated with the disease activity (onset, progression, and remission) in immune-associated conditions which could potentially lead to development of live biotherapeutics.

Hence, the objectives of this research thesis are:

1) Characterisation of the gut microbiota dynamics in various disorders and their subtypes (arthritis, osteoporosis, and multiple sclerosis)
2) To investigate the changes in the microbiota associated with the progression of various arthritic diseases and biologics treatment
3) Development of a novel approach for improved functional inference from 16S microbiota profiles to reflect the bacterial metabolic potential better

8. References

1. *Arthritis* [Online]. Available: https://www.niams.nih.gov/health-topics/arthritis [Accessed].
2. "Taxonomic Outline of the Archaea and Bacteria," in *Bergey's Manual of Systematics of Archaea and Bacteria*.), 1-17.
3. Abdollahi-Roodsaz, S., Joosten, L.A., Koenders, M.I., Devesa, I., Roelofs, M.F., Radstake, T.R., Heuvelmans-Jacobs, M., Akira, S., Nicklin, M.J., Ribeiro-Dias, F., and Van Den Berg, W.B. (2008). Stimulation of TLR2 and TLR4 differentially skews the balance of T cells in a mouse model of arthritis. *J Clin Invest* 118**,** 205-216.
4. Abrahamsson, T.R., Jakobsson, T., Bottcher, M.F., Fredrikson, M., Jenmalm, M.C., Bjorksten, B., and Oldaeus, G. (2007). Probiotics in prevention of IgE-associated eczema: a double-blind, randomized, placebo-controlled trial. *J Allergy Clin Immunol* 119**,** 1174-1180.
5. Abrams, G.D., Bauer, H., and Sprinz, H. (1963). Influence of the normal flora on mucosal morphology and cellular renewal in the ileum. A comparison of germ-free and conventional mice. *Lab Invest* 12**,** 355-364.
6. Abreu, M.T., Fukata, M., and Arditi, M. (2005). TLR signaling in the gut in health and disease. *J Immunol* 174**,** 4453-4460.
7. Abubucker, S., Segata, N., Goll, J., Schubert, A.M., Izard, J., Cantarel, B.L., Rodriguez-Mueller, B., Zucker, J., Thiagarajan, M., Henrissat, B., White, O., Kelley, S.T., Methe, B., Schloss, P.D., Gevers, D., Mitreva, M., and Huttenhower, C. (2012). Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput Biol* 8**,** e1002358.
8. Acinas, S.G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., and Polz, M.F. (2005). PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol* 71**,** 8966-8969.
9. Adegbola, S.O., Sahnan, K., Warusavitarne, J., Hart, A., and Tozer, P. (2018). Anti-TNF Therapy in Crohn's Disease. *Int J Mol Sci* 19.
10. Adlerberth, I., and Wold, A.E. (2009). Establishment of the gut microbiota in Western infants. *Acta Paediatr* 98**,** 229-238.
11. Afiahayati, Sato, K., and Sakakibara, Y. (2015). MetaVelvet-SL: an extension of the Velvet assembler to a de novo metagenomic assembler utilizing supervised learning. *DNA Res* 22**,** 69-77.
12. Afzal, N.A., Davies, S., Paintin, M., Arnaud-Battandier, F., Walker-Smith, J.A., Murch, S., Heuschkel, R., and Fell, J. (2005). Colonic Crohn's disease in children does not respond well to treatment with enteral nutrition if the ileum is not involved. *Dig Dis Sci* 50**,** 1471-1475.
13. Ainsworth, D., Sternberg, M.J.E., Raczy, C., and Butcher, S.A. (2017). k-SLAM: accurate and ultra-fast taxonomic classification and gene identification for large metagenomic data sets. *Nucleic Acids Res* 45**,** 1649-1656.
14. Aletaha, D., Neogi, T., Silman, A.J., Funovits, J., Felson, D.T., Bingham, C.O., 3rd, Birnbaum, N.S., Burmester, G.R., Bykerk, V.P., Cohen, M.D., Combe, B., Costenbader, K.H., Dougados, M., Emery, P., Ferraccioli, G., Hazes, J.M., Hobbs, K., Huizinga, T.W., Kavanaugh, A., Kay, J., Kvien, T.K., Laing, T., Mease, P., Menard, H.A., Moreland, L.W., Naden, R.L., Pincus, T., Smolen, J.S., Stanislawska-Biernat, E., Symmons, D., Tak, P.P., Upchurch, K.S.,

Vencovsky, J., Wolfe, F., and Hawker, G. (2010). 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum* 62**,** 2569-2581.
15. Aletaha, D., and Smolen, J.S. (2007). The Simplified Disease Activity Index (SDAI) and Clinical Disease Activity Index (CDAI) to monitor patients in standard clinical care. *Best Pract Res Clin Rheumatol* 21**,** 663-675.
16. Aletaha, D., and Smolen, J.S. (2018). Diagnosis and Management of Rheumatoid Arthritis: A Review. *JAMA* 320**,** 1360-1372.
17. Ali, T., Kaitha, S., Mahmood, S., Ftesi, A., Stone, J., and Bronze, M.S. (2013). Clinical use of anti-TNF therapy and increased risk of infections. *Drug Healthc Patient Saf* 5**,** 79-99.
18. Alipour, B., Homayouni-Rad, A., Vaghef-Mehrabany, E., Sharif, S.K., Vaghef-Mehrabany, L., Asghari-Jafarabadi, M., Nakhjavani, M.R., and Mohtadi-Nia, J. (2014). Effects of Lactobacillus casei supplementation on disease activity and inflammatory cytokines in rheumatoid arthritis patients: a randomized double-blind clinical trial. *Int J Rheum Dis* 17**,** 519-527.
19. Allard, G., Ryan, F.J., Jeffery, I.B., and Claesson, M.J. (2015). SPINGO: a rapid species-classifier for microbial amplicon sequences. *BMC Bioinformatics* 16**,** 324.
20. Alneberg, J., Bjarnason, B.S., De Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman, N.J., Andersson, A.F., and Quince, C. (2014). Binning metagenomic contigs by coverage and composition. *Nat Methods* 11**,** 1144-1146.
21. Amir, A., Mcdonald, D., Navas-Molina, J.A., Kopylova, E., Morton, J.T., Zech Xu, Z., Kightley, E.P., Thompson, L.R., Hyde, E.R., Gonzalez, A., and Knight, R. (2017). Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *mSystems* 2.
22. Amor, B., Dougados, M., and Mijiyawa, M. (1990). [Criteria of the classification of spondylarthropathies]. *Rev Rhum Mal Osteoartic* 57**,** 85-89.
23. Ananthakrishnan, A.N., Higuchi, L.M., Huang, E.S., Khalili, H., Richter, J.M., Fuchs, C.S., and Chan, A.T. (2012). Aspirin, nonsteroidal anti-inflammatory drug use, and risk for Crohn disease and ulcerative colitis: a cohort study. *Ann Intern Med* 156**,** 350-359.
24. Anderson, C.A., Boucher, G., Lees, C.W., Franke, A., D'amato, M., Taylor, K.D., Lee, J.C., Goyette, P., Imielinski, M., Latiano, A., Lagace, C., Scott, R., Amininejad, L., Bumpstead, S., Baidoo, L., Baldassano, R.N., Barclay, M., Bayless, T.M., Brand, S., Buning, C., Colombel, J.F., Denson, L.A., De Vos, M., Dubinsky, M., Edwards, C., Ellinghaus, D., Fehrmann, R.S., Floyd, J.A., Florin, T., Franchimont, D., Franke, L., Georges, M., Glas, J., Glazer, N.L., Guthery, S.L., Haritunians, T., Hayward, N.K., Hugot, J.P., Jobin, G., Laukens, D., Lawrance, I., Lemann, M., Levine, A., Libioulle, C., Louis, E., Mcgovern, D.P., Milla, M., Montgomery, G.W., Morley, K.I., Mowat, C., Ng, A., Newman, W., Ophoff, R.A., Papi, L., Palmieri, O., Peyrin-Biroulet, L., Panes, J., Phillips, A., Prescott, N.J., Proctor, D.D., Roberts, R., Russell, R., Rutgeerts, P., Sanderson, J., Sans, M., Schumm, P., Seibold, F., Sharma, Y., Simms, L.A., Seielstad, M., Steinhart, A.H., Targan, S.R., Van Den Berg, L.H., Vatn, M., Verspaget, H., Walters, T., Wijmenga, C., Wilson, D.C., Westra, H.J., Xavier, R.J., Zhao, Z.Z., Ponsioen, C.Y., Andersen, V., Torkvist, L., Gazouli, M., Anagnou, N.P., Karlsen, T.H., Kupcinskas, L., Sventoraityte, J., Mansfield, J.C.,

Kugathasan, S., Silverberg, M.S., Halfvarson, J., Rotter, J.I., Mathew, C.G., Griffiths, A.M., Gearry, R., Ahmad, T., Brant, S.R., Chamaillard, M., et al. (2011). Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet* 43**,** 246-252.

25. Andoh, A., Bamba, T., and Sasaki, M. (1999). Physiological and anti-inflammatory roles of dietary fiber and butyrate in intestinal functions. *JPEN J Parenter Enteral Nutr* 23**,** S70-73.
26. Anstadt, E.J., Fujiwara, M., Wasko, N., Nichols, F., and Clark, R.B. (2016). TLR Tolerance as a Treatment for Central Nervous System Autoimmunity. *J Immunol* 197**,** 2110-2118.
27. Appel, H., Maier, R., Bleil, J., Hempfing, A., Loddenkemper, C., Schlichting, U., Syrbe, U., and Sieper, J. (2013). In situ analysis of interleukin-23- and interleukin-12-positive cells in the spine of patients with ankylosing spondylitis. *Arthritis Rheum* 65**,** 1522-1529.
28. Appel, H., Maier, R., Wu, P., Scheer, R., Hempfing, A., Kayser, R., Thiel, A., Radbruch, A., Loddenkemper, C., and Sieper, J. (2011). Analysis of IL-17(+) cells in facet joints of patients with spondyloarthritis suggests that the innate immune pathway might be of greater relevance than the Th17-mediated adaptive immune response. *Arthritis Res Ther* 13**,** R95.
29. Arend, W.P., and Firestein, G.S. (2012). Pre-rheumatoid arthritis: predisposition and transition to clinical synovitis. *Nat Rev Rheumatol* 8**,** 573-586.
30. Armour, C.R., Nayfach, S., Pollard, K.S., and Sharpton, T.J. (2019). A Metagenomic Meta-analysis Reveals Functional Signatures of Health and Disease in the Human Gut Microbiome. *mSystems* 4.
31. Arpaia, N., Campbell, C., Fan, X., Dikiy, S., Van Der Veeken, J., Deroos, P., Liu, H., Cross, J.R., Pfeffer, K., Coffer, P.J., and Rudensky, A.Y. (2013). Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* 504**,** 451-455.
32. Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J.M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H.B., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E.G., Wang, J., Guarner, F., Pedersen, O., De Vos, W.M., Brunak, S., Dore, J., Meta, H.I.T.C., Antolin, M., Artiguenave, F., Blottiere, H.M., Almeida, M., Brechot, C., Cara, C., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Foerstner, K.U., Friss, C., Van De Guchte, M., Guedon, E., Haimet, F., Huber, W., Van Hylckama-Vlieg, J., Jamet, A., Juste, C., Kaci, G., Knol, J., Lakhdari, O., Layec, S., Le Roux, K., Maguin, E., Merieux, A., Melo Minardi, R., M'rini, C., Muller, J., Oozeer, R., Parkhill, J., Renault, P., Rescigno, M., Sanchez, N., Sunagawa, S., Torrejon, A., Turner, K., Vandemeulebrouck, G., Varela, E., Winogradsky, Y., Zeller, G., Weissenbach, J., Ehrlich, S.D., and Bork, P. (2011). Enterotypes of the human gut microbiome. *Nature* 473**,** 174-180.
33. Ash, Z., Gaujoux-Viala, C., Gossec, L., Hensor, E.M., Fitzgerald, O., Winthrop, K., Van Der Heijde, D., Emery, P., Smolen, J.S., and Marzo-Ortega, H. (2012). A systematic literature review of drug therapies for the treatment of psoriatic arthritis: current evidence and meta-analysis informing the EULAR

recommendations for the management of psoriatic arthritis. *Ann Rheum Dis* 71**,** 319-326.
34. Asquith, M., Davin, S., Stauffer, P., Michell, C., Janowitz, C., Lin, P., Ensign-Lewis, J., Kinchen, J.M., Koop, D.R., and Rosenbaum, J.T. (2017). Intestinal Metabolites Are Profoundly Altered in the Context of HLA-B27 Expression and Functionally Modulate Disease in a Rat Model of Spondyloarthritis. *Arthritis Rheumatol* 69**,** 1984-1995.
35. Asquith, M.J., Stauffer, P., Davin, S., Mitchell, C., Lin, P., and Rosenbaum, J.T. (2016). Perturbed Mucosal Immunity and Dysbiosis Accompany Clinical Disease in a Rat Model of Spondyloarthritis. *Arthritis Rheumatol* 68**,** 2151-2162.
36. Asshauer, K.P., Wemheuer, B., Daniel, R., and Meinicke, P. (2015). Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics* 31**,** 2882-2884.
37. Avenell, A., Mak, J.C., and O'connell, D. (2014). Vitamin D and vitamin D analogues for preventing fractures in post-menopausal women and older men. *Cochrane Database Syst Rev***,** CD000227.
38. Azad, M.a.K., Sarker, M., Li, T., and Yin, J. (2018). Probiotic Species in the Modulation of Gut Microbiota: An Overview. *Biomed Res Int* 2018**,** 9478630.
39. Backhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A., and Gordon, J.I. (2005). Host-bacterial mutualism in the human intestine. *Science* 307**,** 1915-1920.
40. Baeten, D., Baraliakos, X., Braun, J., Sieper, J., Emery, P., Van Der Heijde, D., Mcinnes, I., Van Laar, J.M., Landewe, R., Wordsworth, P., Wollenhaupt, J., Kellner, H., Paramarta, J., Wei, J., Brachat, A., Bek, S., Laurent, D., Li, Y., Wang, Y.A., Bertolino, A.P., Gsteiger, S., Wright, A.M., and Hueber, W. (2013). Anti-interleukin-17A monoclonal antibody secukinumab in treatment of ankylosing spondylitis: a randomised, double-blind, placebo-controlled trial. *Lancet* 382**,** 1705-1713.
41. Baldwin, B.G., Sanderson, M.J., Porter, J.M., Wojciechowski, M.F., Campbell, C.S., and Donoghue, M.J. (1995). The its Region of Nuclear Ribosomal DNA: A Valuable Source of Evidence on Angiosperm Phylogeny. *Annals of the Missouri Botanical Garden* 82**,** 247-277.
42. Bang, S., Yoo, D., Kim, S.J., Jhang, S., Cho, S., and Kim, H. (2019). Establishment and evaluation of prediction model for multiple disease classification based on gut microbial data. *Sci Rep* 9**,** 10189.
43. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., and Pevzner, P.A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19**,** 455-477.
44. Baraliakos, X., Haibel, H., Listing, J., Sieper, J., and Braun, J. (2014). Continuous long-term anti-TNF therapy does not lead to an increase in the rate of new bone formation over 8 years in patients with ankylosing spondylitis. *Ann Rheum Dis* 73**,** 710-715.
45. Belizario, J.E., and Faintuch, J. (2018). Microbiome and Gut Dysbiosis. *Exp Suppl* 109**,** 459-476.
46. Belkaid, Y., and Hand, T.W. (2014). Role of the microbiota in immunity and inflammation. *Cell* 157**,** 121-141.

47. Bennike, T., Birkelund, S., Stensballe, A., and Andersen, V. (2014). Biomarkers in inflammatory bowel diseases: current status and proteomics identification strategies. *World J Gastroenterol* 20**,** 3231-3244.
48. Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., Boutell, J.M., Bryant, J., Carter, R.J., Keira Cheetham, R., Cox, A.J., Ellis, D.J., Flatbush, M.R., Gormley, N.A., Humphray, S.J., Irving, L.J., Karbelashvili, M.S., Kirk, S.M., Li, H., Liu, X., Maisinger, K.S., Murray, L.J., Obradovic, B., Ost, T., Parkinson, M.L., Pratt, M.R., Rasolonjatovo, I.M., Reed, M.T., Rigatti, R., Rodighiero, C., Ross, M.T., Sabot, A., Sankar, S.V., Scally, A., Schroth, G.P., Smith, M.E., Smith, V.P., Spiridou, A., Torrance, P.E., Tzonev, S.S., Vermaas, E.H., Walter, K., Wu, X., Zhang, L., Alam, M.D., Anastasi, C., Aniebo, I.C., Bailey, D.M., Bancarz, I.R., Banerjee, S., Barbour, S.G., Baybayan, P.A., Benoit, V.A., Benson, K.F., Bevis, C., Black, P.J., Boodhun, A., Brennan, J.S., Bridgham, J.A., Brown, R.C., Brown, A.A., Buermann, D.H., Bundu, A.A., Burrows, J.C., Carter, N.P., Castillo, N., Chiara, E.C.M., Chang, S., Neil Cooley, R., Crake, N.R., Dada, O.O., Diakoumakos, K.D., Dominguez-Fernandez, B., Earnshaw, D.J., Egbujor, U.C., Elmore, D.W., Etchin, S.S., Ewan, M.R., Fedurco, M., Fraser, L.J., Fuentes Fajardo, K.V., Scott Furey, W., George, D., Gietzen, K.J., Goddard, C.P., Golda, G.S., Granieri, P.A., Green, D.E., Gustafson, D.L., Hansen, N.F., Harnish, K., Haudenschild, C.D., Heyer, N.I., Hims, M.M., Ho, J.T., Horgan, A.M., et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456**,** 53-59.
49. Berer, K., Gerdes, L.A., Cekanaviciute, E., Jia, X., Xiao, L., Xia, Z., Liu, C., Klotz, L., Stauffer, U., Baranzini, S.E., Kumpfel, T., Hohlfeld, R., Krishnamoorthy, G., and Wekerle, H. (2017). Gut microbiota from multiple sclerosis patients enables spontaneous autoimmune encephalomyelitis in mice. *Proc Natl Acad Sci U S A* 114**,** 10719-10724.
50. Berer, K., Mues, M., Koutrolos, M., Rasbi, Z.A., Boziki, M., Johner, C., Wekerle, H., and Krishnamoorthy, G. (2011). Commensal microbiota and myelin autoantigen cooperate to trigger autoimmune demyelination. *Nature* 479**,** 538-541.
51. Bernardo, D., Vallejo-Diez, S., Mann, E.R., Al-Hassi, H.O., Martinez-Abad, B., Montalvillo, E., Tee, C.T., Murugananthan, A.U., Nunez, H., Peake, S.T., Hart, A.L., Fernandez-Salazar, L., Garrote, J.A., Arranz, E., and Knight, S.C. (2012). IL-6 promotes immune responses in human ulcerative colitis and induces a skin-homing phenotype in the dendritic cells and Tcells they stimulate. *Eur J Immunol* 42**,** 1337-1353.
52. Bernstein, H., Bernstein, C., Payne, C.M., Dvorakova, K., and Garewal, H. (2005). Bile acids as carcinogens in human gastrointestinal cancers. *Mutat Res* 589**,** 47-65.
53. Bevilacqua, A., Costabile, A., Bergillos-Meca, T., Gonzalez, I., Landriscina, L., Ciuffreda, E., D'agnello, P., Corbo, M.R., Sinigaglia, M., and Lamacchia, C. (2016). Impact of Gluten-Friendly Bread on the Metabolism and Function of In Vitro Gut Microbiota in Healthy Human and Coeliac Subjects. *PLoS One* 11**,** e0162770.
54. Bian, X., Chi, L., Gao, B., Tu, P., Ru, H., and Lu, K. (2017). Gut Microbiome Response to Sucralose and Its Potential Role in Inducing Liver Inflammation in Mice. *Front Physiol* 8**,** 487.

55. Bindels, L.B., Delzenne, N.M., Cani, P.D., and Walter, J. (2015). Towards a more comprehensive concept for prebiotics. *Nat Rev Gastroenterol Hepatol* 12**,** 303-310.
56. Bindels, L.B., Porporato, P., Dewulf, E.M., Verrax, J., Neyrinck, A.M., Martin, J.C., Scott, K.P., Buc Calderon, P., Feron, O., Muccioli, G.G., Sonveaux, P., Cani, P.D., and Delzenne, N.M. (2012). Gut microbiota-derived propionate reduces cancer cell proliferation in the liver. *Br J Cancer* 107**,** 1337-1344.
57. Biniecka, M., Canavan, M., Mcgarry, T., Gao, W., Mccormick, J., Cregan, S., Gallagher, L., Smith, T., Phelan, J.J., Ryan, J., O'sullivan, J., Ng, C.T., Veale, D.J., and Fearon, U. (2016). Dysregulated bioenergetics: a key regulator of joint inflammation. *Ann Rheum Dis* 75**,** 2192-2200.
58. Biniecka, M., Kennedy, A., Fearon, U., Ng, C.T., Veale, D.J., and O'sullivan, J.N. (2010). Oxidative damage in synovial tissue is associated with in vivo hypoxic status in the arthritic joint. *Ann Rheum Dis* 69**,** 1172-1178.
59. Biton, M., Haber, A.L., Rogel, N., Burgin, G., Beyaz, S., Schnell, A., Ashenberg, O., Su, C.W., Smillie, C., Shekhar, K., Chen, Z., Wu, C., Ordovas-Montanes, J., Alvarez, D., Herbst, R.H., Zhang, M., Tirosh, I., Dionne, D., Nguyen, L.T., Xifaras, M.E., Shalek, A.K., Von Andrian, U.H., Graham, D.B., Rozenblatt-Rosen, O., Shi, H.N., Kuchroo, V., Yilmaz, O.H., Regev, A., and Xavier, R.J. (2018). T Helper Cell Cytokines Modulate Intestinal Stem Cell Renewal and Differentiation. *Cell* 175**,** 1307-1320 e1322.
60. Blanton, L.V., Charbonneau, M.R., Salih, T., Barratt, M.J., Venkatesh, S., Ilkaveya, O., Subramanian, S., Manary, M.J., Trehan, I., Jorgensen, J.M., Fan, Y.M., Henrissat, B., Leyn, S.A., Rodionov, D.A., Osterman, A.L., Maleta, K.M., Newgard, C.B., Ashorn, P., Dewey, K.G., and Gordon, J.I. (2016). Gut bacteria that prevent growth impairments transmitted by microbiota from malnourished children. *Science* 351.
61. Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R., and Abebe, E. (2005). Defining operational taxonomic units using DNA barcode data. *Philos Trans R Soc Lond B Biol Sci* 360**,** 1935-1943.
62. Bluml, S., Redlich, K., and Smolen, J.S. (2014). Mechanisms of tissue damage in arthritis. *Semin Immunopathol* 36**,** 531-540.
63. Bohmig, G.A., Krieger, P.M., Saemann, M.D., Wenhardt, C., Pohanka, E., and Zlabinger, G.J. (1997). n-butyrate downregulates the stimulatory function of peripheral blood-derived antigen-presenting cells: a potential mechanism for modulating T-cell responses by short-chain fatty acids. *Immunology* 92**,** 234-243.
64. Bokulich, N.A., Chung, J., Battaglia, T., Henderson, N., Jay, M., Li, H., A, D.L., Wu, F., Perez-Perez, G.I., Chen, Y., Schweizer, W., Zheng, X., Contreras, M., Dominguez-Bello, M.G., and Blaser, M.J. (2016). Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Sci Transl Med* 8**,** 343ra382.
65. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30**,** 2114-2120.
66. Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J.E., Bittinger, K., Brejnrod, A., Brislawn, C.J., Brown, C.T., Callahan, B.J., Caraballo-Rodriguez, A.M., Chase, J., Cope, E.K., Da Silva, R., Diener, C., Dorrestein, P.C., Douglas, G.M., Durall, D.M., Duvallet, C., Edwardson, C.F., Ernst, M., Estaki, M., Fouquier, J., Gauglitz, J.M., Gibbons, S.M., Gibson,

D.L., Gonzalez, A., Gorlick, K., Guo, J., Hillmann, B., Holmes, S., Holste, H., Huttenhower, C., Huttley, G.A., Janssen, S., Jarmusch, A.K., Jiang, L., Kaehler, B.D., Kang, K.B., Keefe, C.R., Keim, P., Kelley, S.T., Knights, D., Koester, I., Kosciolek, T., Kreps, J., Langille, M.G.I., Lee, J., Ley, R., Liu, Y.X., Loftfield, E., Lozupone, C., Maher, M., Marotz, C., Martin, B.D., Mcdonald, D., Mciver, L.J., Melnik, A.V., Metcalf, J.L., Morgan, S.C., Morton, J.T., Naimey, A.T., Navas-Molina, J.A., Nothias, L.F., Orchanian, S.B., Pearson, T., Peoples, S.L., Petras, D., Preuss, M.L., Pruesse, E., Rasmussen, L.B., Rivers, A., Robeson, M.S., 2nd, Rosenthal, P., Segata, N., Shaffer, M., Shiffer, A., Sinha, R., Song, S.J., Spear, J.R., Swafford, A.D., Thompson, L.R., Torres, P.J., Trinh, P., Tripathi, A., Turnbaugh, P.J., Ul-Hasan, S., Van Der Hooft, J.J.J., Vargas, F., Vazquez-Baeza, Y., Vogtmann, E., Von Hippel, M., Walters, W., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 37**,** 852-857.

67. Bowness, P. (2015). Hla-B27. *Annu Rev Immunol* 33**,** 29-48.
68. Branton, D., Deamer, D.W., Marziali, A., Bayley, H., Benner, S.A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., Jovanovich, S.B., Krstic, P.S., Lindsay, S., Ling, X.S., Mastrangelo, C.H., Meller, A., Oliver, J.S., Pershin, Y.V., Ramsey, J.M., Riehn, R., Soni, G.V., Tabard-Cossa, V., Wanunu, M., Wiggin, M., and Schloss, J.A. (2008). The potential and challenges of nanopore sequencing. *Nat Biotechnol* 26**,** 1146-1153.
69. Braun, J., Baraliakos, X., Golder, W., Brandt, J., Rudwaleit, M., Listing, J., Bollow, M., Sieper, J., and Van Der Heijde, D. (2003). Magnetic resonance imaging examinations of the spine in patients with ankylosing spondylitis, before and after successful therapy with infliximab: evaluation of a new scoring system. *Arthritis Rheum* 48**,** 1126-1136.
70. Braun, J., Bollow, M., Neure, L., Seipelt, E., Seyrekbasan, F., Herbst, H., Eggens, U., Distler, A., and Sieper, J. (1995). Use of immunohistologic and in situ hybridization techniques in the examination of sacroiliac joint biopsy specimens from patients with ankylosing spondylitis. *Arthritis Rheum* 38**,** 499-505.
71. Braun, J., Zochling, J., Baraliakos, X., Alten, R., Burmester, G., Grasedyck, K., Brandt, J., Haibel, H., Hammer, M., Krause, A., Mielke, F., Tony, H.P., Ebner, W., Gomor, B., Hermann, J., Zeidler, H., Beck, E., Baumgaertner, M., and Sieper, J. (2006). Efficacy of sulfasalazine in patients with inflammatory back pain due to undifferentiated spondyloarthritis and early ankylosing spondylitis: a multicentre randomised controlled trial. *Ann Rheum Dis* 65**,** 1147-1153.
72. Bray, J.R., and Curtis, J.T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs* 27**,** 325-349.
73. Breese, E., Braegger, C.P., Corrigan, C.J., Walker-Smith, J.A., and Macdonald, T.T. (1993). Interleukin-2- and interferon-gamma-secreting T cells in normal and diseased human intestinal mucosa. *Immunology* 78**,** 127-131.
74. Breitwieser, F.P., Lu, J., and Salzberg, S.L. (2017). A review of methods and databases for metagenomic classification and assembly. *Brief Bioinform*.
75. Brewerton, D.A., Hart, F.D., Nicholls, A., Caffrey, M., James, D.C., and Sturrock, R.D. (1973). Ankylosing spondylitis and HL-A 27. *Lancet* 1**,** 904-907.
76. Brown, E.M., Kenny, D.J., and Xavier, R.J. (2019). Gut Microbiota Regulation of T Cells During Inflammation and Autoimmunity. *Annu Rev Immunol* 37**,** 599-624.

77. Brown, M.A., Laval, S.H., Brophy, S., and Calin, A. (2000). Recurrence risk modelling of the genetic susceptibility to ankylosing spondylitis. *Ann Rheum Dis* 59**,** 883-886.
78. Brynedal, B., Duvefelt, K., Jonasdottir, G., Roos, I.M., Akesson, E., Palmgren, J., and Hillert, J. (2007). HLA-A confers an HLA-DRB1 independent influence on the risk of multiple sclerosis. *PLoS One* 2**,** e664.
79. Buc, M. (2018). New biological agents in the treatment of multiple sclerosis. *Bratisl Lek Listy* 119**,** 191-197.
80. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12**,** 59-60.
81. Bullard, J.H., Purdom, E., Hansen, K.D., and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* 11**,** 94.
82. Byndloss, M.X., Olsan, E.E., Rivera-Chavez, F., Tiffany, C.R., Cevallos, S.A., Lokken, K.L., Torres, T.P., Byndloss, A.J., Faber, F., Gao, Y., Litvak, Y., Lopez, C.A., Xu, G., Napoli, E., Giulivi, C., Tsolis, R.M., Revzin, A., Lebrilla, C.B., and Baumler, A.J. (2017). Microbiota-activated PPAR-gamma signaling inhibits dysbiotic Enterobacteriaceae expansion. *Science* 357**,** 570-575.
83. Callahan, B.J., Mcmurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 13**,** 581-583.
84. Candia, L., Marquez, J., Hernandez, C., Zea, A.H., and Espinoza, L.R. (2007). Toll-like receptor-2 expression is upregulated in antigen-presenting cells from patients with psoriatic arthritis: a pathogenic role for innate immunity? *J Rheumatol* 34**,** 374-379.
85. Cantini, F., Niccoli, L., Nannini, C., Cassara, E., Kaloudi, O., Giulio Favalli, E., Becciolini, A., Biggioggero, M., Benucci, M., Li Gobbi, F., Grossi, V., Infantino, M., Meacci, F., Manfredi, M., Guiducci, S., Bellando-Randone, S., Matucci-Cerinic, M., Foti, R., Di Gangi, M., Mosca, M., Tani, C., Palmieri, F., Goletti, D., and Italian Board for The, T.B.T. (2016). Tailored first-line biologic therapy in patients with rheumatoid arthritis, spondyloarthritis, and psoriatic arthritis. *Semin Arthritis Rheum* 45**,** 519-532.
86. Cantorna, M.T. (2006). Vitamin D and its role in immunology: multiple sclerosis, and inflammatory bowel disease. *Prog Biophys Mol Biol* 92**,** 60-64.
87. Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., Mcdonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., and Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7**,** 335-336.
88. Carneiro, S., Azevedo, V.F., Bonfiglioli, R., Ranza, R., Goncalves, C.R., Keiserman, M., Meirelles Ede, S., Pinheiro Mde, M., Ximenes, A.C., Bernardo, W., Sampaio-Barros, P.D., and Comissao De Espondiloartrites Da Sociedade Brasileira De, R. (2013). Recommendations for the management and treatment of psoriatic arthritis. *Rev Bras Reumatol* 53**,** 227-241.
89. Carter, M.J., Lobo, A.J., Travis, S.P., and Ibd Section, B.S.O.G. (2004). Guidelines for the management of inflammatory bowel disease in adults. *Gut* 53 Suppl 5**,** V1-16.

90. Ceponis, A., Konttinen, Y.T., Imai, S., Tamulaitiene, M., Li, T.F., Xu, J.W., Hietanen, J., Santavirta, S., and Fassbender, H.G. (1998). Synovial lining, endothelial and inflammatory mononuclear cell proliferation in synovial membranes in psoriatic and reactive arthritis: a comparative quantitative morphometric study. *Br J Rheumatol* 37**,** 170-178.
91. Chai, J.N., Peng, Y., Rengarajan, S., Solomon, B.D., Ai, T.L., Shen, Z., Perry, J.S.A., Knoop, K.A., Tanoue, T., Narushima, S., Honda, K., Elson, C.O., Newberry, R.D., Stappenbeck, T.S., Kau, A.L., Peterson, D.A., Fox, J.G., and Hsieh, C.S. (2017). Helicobacter species are potent drivers of colonic T cell responses in homeostasis and inflammation. *Sci Immunol* 2.
92. Chakravorty, S., Helb, D., Burday, M., Connell, N., and Alland, D. (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods* 69**,** 330-339.
93. Chao, A., Chazdon, R.L., Colwell, R.K., and Shen, T.J. (2006). Abundance-based similarity indices and their estimation when there are unseen species in samples. *Biometrics* 62**,** 361-371.
94. Chassaing, B., Koren, O., Goodrich, J.K., Poole, A.C., Srinivasan, S., Ley, R.E., and Gewirtz, A.T. (2015). Dietary emulsifiers impact the mouse gut microbiota promoting colitis and metabolic syndrome. *Nature* 519**,** 92-96.
95. Chen, J., Wright, K., Davis, J.M., Jeraldo, P., Marietta, E.V., Murray, J., Nelson, H., Matteson, E.L., and Taneja, V. (2016). An expansion of rare lineage intestinal microbes characterizes rheumatoid arthritis. *Genome Med* 8**,** 43.
96. Chen, L., Deng, H., Cui, H., Fang, J., Zuo, Z., Deng, J., Li, Y., Wang, X., and Zhao, L. (2018). Inflammatory responses and inflammation-associated diseases in organs. *Oncotarget* 9**,** 7204-7218.
97. Chu, F., Shi, M., Lang, Y., Shen, D., Jin, T., Zhu, J., and Cui, L. (2018). Gut Microbiota in Multiple Sclerosis and Experimental Autoimmune Encephalomyelitis: Current Applications and Future Perspectives. *Mediators Inflamm* 2018**,** 8168717.
98. Claesson, M.J., Clooney, A.G., and O'toole, P.W. (2017). A clinician's guide to microbiome analysis. *Nat Rev Gastroenterol Hepatol* 14**,** 585-595.
99. Claesson, M.J., Jeffery, I.B., Conde, S., Power, S.E., O'connor, E.M., Cusack, S., Harris, H.M., Coakley, M., Lakshminarayanan, B., O'sullivan, O., Fitzgerald, G.F., Deane, J., O'connor, M., Harnedy, N., O'connor, K., O'mahony, D., Van Sinderen, D., Wallace, M., Brennan, L., Stanton, C., Marchesi, J.R., Fitzgerald, A.P., Shanahan, F., Hill, C., Ross, R.P., and O'toole, P.W. (2012). Gut microbiota composition correlates with diet and health in the elderly. *Nature* 488**,** 178-184.
100. Clemente, J.C., Manasson, J., and Scher, J.U. (2018). The role of the gut microbiome in systemic inflammatory disease. *BMJ* 360**,** j5145.
101. Clemente, J.C., Pehrsson, E.C., Blaser, M.J., Sandhu, K., Gao, Z., Wang, B., Magris, M., Hidalgo, G., Contreras, M., Noya-Alarcon, O., Lander, O., Mcdonald, J., Cox, M., Walter, J., Oh, P.L., Ruiz, J.F., Rodriguez, S., Shen, N., Song, S.J., Metcalf, J., Knight, R., Dantas, G., and Dominguez-Bello, M.G. (2015). The microbiome of uncontacted Amerindians. *Sci Adv* 1.
102. Clooney, A.G., Bernstein, C.N., Leslie, W.D., Vagianos, K., Sargent, M., Laserna-Mendieta, E.J., Claesson, M.J., and Targownik, L.E. (2016). A comparison of the gut microbiome between long-term users and non-users of proton pump inhibitors. *Aliment Pharmacol Ther* 43**,** 974-984.

103. Coates, L.C., Tillett, W., Chandler, D., Helliwell, P.S., Korendowych, E., Kyle, S., Mcinnes, I.B., Oliver, S., Ormerod, A., Smith, C., Symmons, D., Waldron, N., Mchugh, N.J., Committee, B.S.R.C.A., Standards, A., Guidelines Working, G., and The, B. (2013). The 2012 BSR and BHPR guideline for the treatment of psoriatic arthritis with biologics. *Rheumatology (Oxford)* 52**,** 1754-1757.
104. Codoner, F.M., Ramirez-Bosca, A., Climent, E., Carrion-Gutierrez, M., Guerrero, M., Perez-Orquin, J.M., Horga De La Parte, J., Genoves, S., Ramon, D., Navarro-Lopez, V., and Chenoll, E. (2018). Gut microbial composition in patients with psoriasis. *Sci Rep* 8**,** 3812.
105. Coffre, M., Roumier, M., Rybczynska, M., Sechet, E., Law, H.K., Gossec, L., Dougados, M., Bianchi, E., and Rogge, L. (2013). Combinatorial control of Th17 and Th1 cell functions by genetic variations in genes associated with the interleukin-23 signaling pathway in spondyloarthritis. *Arthritis Rheum* 65**,** 1510-1521.
106. Cojocaru, M., Cojocaru, I.M., Silosi, I., Vrabie, C.D., and Tanasescu, R. (2010). Extra-articular Manifestations in Rheumatoid Arthritis. *Maedica (Buchar)* 5**,** 286-291.
107. Cole, J.R., Wang, Q., Fish, J.A., Chai, B., Mcgarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., and Tiedje, J.M. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 42**,** D633-642.
108. Collaborators, G.B.D.M.S. (2019). Global, regional, and national burden of multiple sclerosis 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol* 18**,** 269-285.
109. Comi, G., Radaelli, M., and Soelberg Sorensen, P. (2017). Evolving concepts in the treatment of relapsing multiple sclerosis. *Lancet* 389**,** 1347-1356.
110. Compston, A., and Coles, A. (2002). Multiple sclerosis. *Lancet* 359**,** 1221-1231.
111. Compston, A., and Coles, A. (2008). Multiple sclerosis. *Lancet* 372**,** 1502-1517.
112. Cong, Y., Feng, T., Fujihashi, K., Schoeb, T.R., and Elson, C.O. (2009). A dominant, coordinated T regulatory cell-IgA response to the intestinal microbiota. *Proc Natl Acad Sci U S A* 106**,** 19256-19261.
113. Costello, M.E., Ciccia, F., Willner, D., Warrington, N., Robinson, P.C., Gardiner, B., Marshall, M., Kenna, T.J., Triolo, G., and Brown, M.A. (2015). Brief Report: Intestinal Dysbiosis in Ankylosing Spondylitis. *Arthritis Rheumatol* 67**,** 686-691.
114. Crowson, C.S., Liao, K.P., Davis, J.M., 3rd, Solomon, D.H., Matteson, E.L., Knutson, K.L., Hlatky, M.A., and Gabriel, S.E. (2013). Rheumatoid arthritis and cardiovascular disease. *Am Heart J* 166**,** 622-628 e621.
115. Cuchacovich, R., Perez-Alamino, R., Garcia-Valladares, I., and Espinoza, L.R. (2012). Steps in the management of psoriatic arthritis: a guide for clinicians. *Ther Adv Chronic Dis* 3**,** 259-269.
116. Cussotto, S., Clarke, G., Dinan, T.G., and Cryan, J.F. (2019). Psychotropics and the Microbiome: a Chamber of Secrets. *Psychopharmacology (Berl)* 236**,** 1411-1432.
117. Czesnikiewicz-Guzik, M., Lee, W.W., Cui, D., Hiruma, Y., Lamar, D.L., Yang, Z.Z., Ouslander, J.G., Weyand, C.M., and Goronzy, J.J. (2008). T cell subset-specific susceptibility to aging. *Clin Immunol* 127**,** 107-118.

118. Dabritz, J., Gerner, P., Enninger, A., Classen, M., and Radke, M. (2017). Inflammatory Bowel Disease in Childhood and Adolescence. *Dtsch Arztebl Int* 114**,** 331-338.
119. Das, B., Ghosh, T.S., Kedia, S., Rampal, R., Saxena, S., Bag, S., Mitra, R., Dayal, M., Mehta, O., Surendranath, A., Travis, S.P.L., Tripathi, P., Nair, G.B., and Ahuja, V. (2018). Analysis of the Gut Microbiome of Rural and Urban Healthy Indians Living in Sea Level and High Altitude Areas. *Sci Rep* 8**,** 10104.
120. Davis, C., Kota, K., Baldhandapani, V., Gong, W., Abubucker, S., Becker, E., Martin, J., Wylie, K.M., Khetani, R., Hudson, M.E., Weinstock, G.M., and Mitreva, M. (2015). mBLAST: Keeping up with the sequencing explosion for (meta)genome analysis. *J Data Mining Genomics Proteomics* 4.
121. De Goffau, M.C., Lager, S., Sovio, U., Gaccioli, F., Cook, E., Peacock, S.J., Parkhill, J., Charnock-Jones, D.S., and Smith, G.C.S. (2019). Human placenta has no microbiome but can contain potential pathogens. *Nature* 572**,** 329-334.
122. De Jager, P.L., Jia, X., Wang, J., De Bakker, P.I., Ottoboni, L., Aggarwal, N.T., Piccio, L., Raychaudhuri, S., Tran, D., Aubin, C., Briskin, R., Romano, S., International, M.S.G.C., Baranzini, S.E., Mccauley, J.L., Pericak-Vance, M.A., Haines, J.L., Gibson, R.A., Naeglin, Y., Uitdehaag, B., Matthews, P.M., Kappos, L., Polman, C., Mcardle, W.L., Strachan, D.P., Evans, D., Cross, A.H., Daly, M.J., Compston, A., Sawcer, S.J., Weiner, H.L., Hauser, S.L., Hafler, D.A., and Oksenberg, J.R. (2009). Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. *Nat Genet* 41**,** 776-782.
123. De La Cuesta-Zuluaga, J., Mueller, N.T., Corrales-Agudelo, V., Velasquez-Mejia, E.P., Carmona, J.A., Abad, J.M., and Escobar, J.S. (2017). Metformin Is Associated With Higher Relative Abundance of Mucin-Degrading Akkermansia muciniphila and Several Short-Chain Fatty Acid-Producing Microbiota in the Gut. *Diabetes Care* 40**,** 54-62.
124. De Vos, M., Cuvelier, C., Mielants, H., Veys, E., Barbier, F., and Elewaut, A. (1989). Ileocolonoscopy in seronegative spondylarthropathy. *Gastroenterology* 96**,** 339-344.
125. Dean, L.E., Jones, G.T., Macdonald, A.G., Downham, C., Sturrock, R.D., and Macfarlane, G.J. (2014). Global prevalence of ankylosing spondylitis. *Rheumatology (Oxford)* 53**,** 650-657.
126. Del Rincon, I., Battafarano, D.F., Restrepo, J.F., Erikson, J.M., and Escalante, A. (2014). Glucocorticoid dose thresholds associated with all-cause and cardiovascular mortality in rheumatoid arthritis. *Arthritis Rheumatol* 66**,** 264-272.
127. Derrien, M., Vaughan, E.E., Plugge, C.M., and De Vos, W.M. (2004). Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int J Syst Evol Microbiol* 54**,** 1469-1476.
128. Desantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G.L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72**,** 5069-5072.
129. Di Sabatino, A., Biancheri, P., Rovedatti, L., Macdonald, T.T., and Corazza, G.R. (2012). New pathogenic paradigms in inflammatory bowel disease. *Inflamm Bowel Dis* 18**,** 368-371.

130. Diez-Perez, A., Adachi, J.D., Agnusdei, D., Bilezikian, J.P., Compston, J.E., Cummings, S.R., Eastell, R., Eriksen, E.F., Gonzalez-Macias, J., Liberman, U.A., Wahl, D.A., Seeman, E., Kanis, J.A., Cooper, C., and Group, I.C.I.R.W. (2012). Treatment failure in osteoporosis. *Osteoporos Int* 23**,** 2769-2774.
131. Dill-Mcfarland, K.A., Tang, Z.Z., Kemis, J.H., Kerby, R.L., Chen, G., Palloni, A., Sorenson, T., Rey, F.E., and Herd, P. (2019). Close social relationships correlate with human gut microbiota composition. *Sci Rep* 9**,** 703.
132. Dominguez-Bello, M.G., Costello, E.K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., and Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci U S A* 107**,** 11971-11975.
133. Dougados, M., Van Der Linden, S., Juhlin, R., Huitfeldt, B., Amor, B., Calin, A., Cats, A., Dijkmans, B., Olivieri, I., Pasero, G., and Et Al. (1991). The European Spondylarthropathy Study Group preliminary criteria for the classification of spondylarthropathy. *Arthritis Rheum* 34**,** 1218-1227.
134. Duffin, K.C., Freeny, I.C., Schrodi, S.J., Wong, B., Feng, B.J., Soltani-Arabshahi, R., Rakkhit, T., Goldgar, D.E., and Krueger, G.G. (2009). Association between IL13 polymorphisms and psoriatic arthritis is modified by smoking. *J Invest Dermatol* 129**,** 2777-2783.
135. Duvallet, C., Gibbons, S.M., Gurry, T., Irizarry, R.A., and Alm, E.J. (2017). Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat Commun* 8**,** 1784.
136. Eastell, R. (1998). Treatment of postmenopausal osteoporosis. *N Engl J Med* 338**,** 736-746.
137. Eastell, R., O'neill, T.W., Hofbauer, L.C., Langdahl, B., Reid, I.R., Gold, D.T., and Cummings, S.R. (2016). Postmenopausal osteoporosis. *Nat Rev Dis Primers* 2**,** 16069.
138. Eder, L., Chandran, V., Pellett, F., Pollock, R., Shanmugarajah, S., Rosen, C.F., Rahman, P., and Gladman, D.D. (2011). IL13 gene polymorphism is a marker for psoriatic arthritis among psoriasis patients. *Ann Rheum Dis* 70**,** 1594-1598.
139. Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26**,** 2460-2461.
140. Edgar, R.C. (2016). UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *bioRxiv***,** 081257.
141. Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27**,** 2194-2200.
142. Edger, R.C. "UTAX accurately predicts taxonomy of marker gene sequences".).
143. Ege, M.J., Mayer, M., Normand, A.C., Genuneit, J., Cookson, W.O., Braun-Fahrlander, C., Heederik, D., Piarroux, R., Von Mutius, E., and Group, G.T.S. (2011). Exposure to environmental microorganisms and childhood asthma. *N Engl J Med* 364**,** 701-709.
144. Egerer, K., Feist, E., and Burmester, G.R. (2009). The serological diagnosis of rheumatoid arthritis: antibodies to citrullinated antigens. *Dtsch Arztebl Int* 106**,** 159-163.
145. Eiseman, B., Silen, W., Bascom, G.S., and Kauvar, A.J. (1958). Fecal enema as an adjunct in the treatment of pseudomembranous enterocolitis. *Surgery* 44**,** 854-859.

146. Eisenhofer, R., Minich, J.J., Marotz, C., Cooper, A., Knight, R., and Weyrich, L.S. (2019). Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. *Trends Microbiol* 27**,** 105-117.
147. Ek, W.E., D'amato, M., and Halfvarson, J. (2014). The history of genetics in inflammatory bowel disease. *Ann Gastroenterol* 27**,** 294-303.
148. Elders, P.J., Lips, P., Netelenbos, J.C., Van Ginkel, F.C., Khoe, E., Van Der Vijgh, W.J., and Van Der Stelt, P.F. (1994). Long-term effect of calcium supplementation on bone loss in perimenopausal women. *J Bone Miner Res* 9**,** 963-970.
149. Eng, A., and Borenstein, E. (2018). Taxa-function robustness in microbial communities. *Microbiome* 6**,** 45.
150. Eren, A.M., Maignien, L., Sul, W.J., Murphy, L.G., Grim, S.L., Morrison, H.G., and Sogin, M.L. (2013). Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol Evol* 4.
151. Eren, A.M., Morrison, H.G., Lescault, P.J., Reveillaud, J., Vineis, J.H., and Sogin, M.L. (2015). Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J* 9**,** 968-979.
152. Evans, D.M., Spencer, C.C., Pointon, J.J., Su, Z., Harvey, D., Kochan, G., Oppermann, U., Dilthey, A., Pirinen, M., Stone, M.A., Appleton, L., Moutsianas, L., Leslie, S., Wordsworth, T., Kenna, T.J., Karaderi, T., Thomas, G.P., Ward, M.M., Weisman, M.H., Farrar, C., Bradbury, L.A., Danoy, P., Inman, R.D., Maksymowych, W., Gladman, D., Rahman, P., Spondyloarthritis Research Consortium Of, C., Morgan, A., Marzo-Ortega, H., Bowness, P., Gaffney, K., Gaston, J.S., Smith, M., Bruges-Armas, J., Couto, A.R., Sorrentino, R., Paladini, F., Ferreira, M.A., Xu, H., Liu, Y., Jiang, L., Lopez-Larrea, C., Diaz-Pena, R., Lopez-Vazquez, A., Zayats, T., Band, G., Bellenguez, C., Blackburn, H., Blackwell, J.M., Bramon, E., Bumpstead, S.J., Casas, J.P., Corvin, A., Craddock, N., Deloukas, P., Dronov, S., Duncanson, A., Edkins, S., Freeman, C., Gillman, M., Gray, E., Gwilliam, R., Hammond, N., Hunt, S.E., Jankowski, J., Jayakumar, A., Langford, C., Liddle, J., Markus, H.S., Mathew, C.G., Mccann, O.T., Mccarthy, M.I., Palmer, C.N., Peltonen, L., Plomin, R., Potter, S.C., Rautanen, A., Ravindrarajah, R., Ricketts, M., Samani, N., Sawcer, S.J., Strange, A., Trembath, R.C., Viswanathan, A.C., Waller, M., Weston, P., Whittaker, P., Widaa, S., Wood, N.W., Mcvean, G., Reveille, J.D., Wordsworth, B.P., Brown, M.A., Donnelly, P., Australo-Anglo-American Spondyloarthritis, C., and Wellcome Trust Case Control, C. (2011). Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nat Genet* 43**,** 761-767.
153. Everard, A., Belzer, C., Geurts, L., Ouwerkerk, J.P., Druart, C., Bindels, L.B., Guiot, Y., Derrien, M., Muccioli, G.G., Delzenne, N.M., De Vos, W.M., and Cani, P.D. (2013). Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. *Proc Natl Acad Sci U S A* 110**,** 9066-9071.
154. Faith, D.P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation* 61**,** 1-10.
155. Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K., Kurilshikov, A., Bonder, M.J., Valles-Colomer, M., Vandeputte, D., Tito, R.Y., Chaffron, S., Rymenans, L., Verspecht, C., De Sutter, L., Lima-Mendez, G.,

D'hoe, K., Jonckheere, K., Homola, D., Garcia, R., Tigchelaar, E.F., Eeckhaudt, L., Fu, J., Henckaerts, L., Zhernakova, A., Wijmenga, C., and Raes, J. (2016). Population-level analysis of gut microbiome variation. *Science* 352**,** 560-564.

156. Fanaro, S., Chierici, R., Guerrini, P., and Vigi, V. (2003). Intestinal microflora in early infancy: composition and development. *Acta Paediatr Suppl* 91**,** 48-55.
157. Fearon, U., Griosios, K., Fraser, A., Reece, R., Emery, P., Jones, P.F., and Veale, D.J. (2003). Angiopoietins, growth factors, and vascular morphology in early arthritis. *J Rheumatol* 30**,** 260-268.
158. Feld, J., Chandran, V., Haroon, N., Inman, R., and Gladman, D. (2018). Axial disease in psoriatic arthritis and ankylosing spondylitis: a critical comparison. *Nat Rev Rheumatol* 14**,** 363-371.
159. Feldtkeller, E., Khan, M.A., Van Der Heijde, D., Van Der Linden, S., and Braun, J. (2003). Age at disease onset and diagnosis delay in HLA-B27 negative vs. positive patients with ankylosing spondylitis. *Rheumatol Int* 23**,** 61-66.
160. Felson, D.T., Smolen, J.S., Wells, G., Zhang, B., Van Tuyl, L.H., Funovits, J., Aletaha, D., Allaart, C.F., Bathon, J., Bombardieri, S., Brooks, P., Brown, A., Matucci-Cerinic, M., Choi, H., Combe, B., De Wit, M., Dougados, M., Emery, P., Furst, D., Gomez-Reino, J., Hawker, G., Keystone, E., Khanna, D., Kirwan, J., Kvien, T.K., Landewe, R., Listing, J., Michaud, K., Martin-Mola, E., Montie, P., Pincus, T., Richards, P., Siegel, J.N., Simon, L.S., Sokka, T., Strand, V., Tugwell, P., Tyndall, A., Van Der Heijde, D., Verstappen, S., White, B., Wolfe, F., Zink, A., and Boers, M. (2011). American College of Rheumatology/European League against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. *Ann Rheum Dis* 70**,** 404-413.
161. Fernandes, A.D., Macklaim, J.M., Linn, T.G., Reid, G., and Gloor, G.B. (2013). ANOVA-like differential expression (ALDEx) analysis for mixed population RNA-Seq. *PLoS One* 8**,** e67019.
162. Ferretti, P., Pasolli, E., Tett, A., Asnicar, F., Gorfer, V., Fedi, S., Armanini, F., Truong, D.T., Manara, S., Zolfo, M., Beghini, F., Bertorelli, R., De Sanctis, V., Bariletti, I., Canto, R., Clementi, R., Cologna, M., Crifo, T., Cusumano, G., Gottardi, S., Innamorati, C., Mase, C., Postai, D., Savoi, D., Duranti, S., Lugli, G.A., Mancabelli, L., Turroni, F., Ferrario, C., Milani, C., Mangifesta, M., Anzalone, R., Viappiani, A., Yassour, M., Vlamakis, H., Xavier, R., Collado, C.M., Koren, O., Tateo, S., Soffiati, M., Pedrotti, A., Ventura, M., Huttenhower, C., Bork, P., and Segata, N. (2018). Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome. *Cell Host Microbe* 24**,** 133-145 e135.
163. Fleischmann, R., Mysler, E., Hall, S., Kivitz, A.J., Moots, R.J., Luo, Z., Demasi, R., Soma, K., Zhang, R., Takiya, L., Tatulych, S., Mojcik, C., Krishnaswami, S., Menon, S., Smolen, J.S., and Investigators, O.S. (2017). Efficacy and safety of tofacitinib monotherapy, tofacitinib with methotrexate, and adalimumab with methotrexate in patients with rheumatoid arthritis (ORAL Strategy): a phase 3b/4, double-blind, head-to-head, randomised controlled trial. *Lancet* 390**,** 457-468.
164. Forbes, J.D., Chen, C.Y., Knox, N.C., Marrie, R.A., El-Gabalawy, H., De Kievit, T., Alfa, M., Bernstein, C.N., and Van Domselaar, G. (2018). A comparative study of the gut microbiota in immune-mediated inflammatory diseases-does a common dysbiosis exist? *Microbiome* 6**,** 221.

165. Forslund, K., Hildebrand, F., Nielsen, T., Falony, G., Le Chatelier, E., Sunagawa, S., Prifti, E., Vieira-Silva, S., Gudmundsdottir, V., Pedersen, H.K., Arumugam, M., Kristiansen, K., Voigt, A.Y., Vestergaard, H., Hercog, R., Costea, P.I., Kultima, J.R., Li, J., Jorgensen, T., Levenez, F., Dore, J., Meta, H.I.T.C., Nielsen, H.B., Brunak, S., Raes, J., Hansen, T., Wang, J., Ehrlich, S.D., Bork, P., and Pedersen, O. (2015). Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* 528**,** 262-266.
166. Franceschi, C., Capri, M., Monti, D., Giunta, S., Olivieri, F., Sevini, F., Panourgia, M.P., Invidia, L., Celani, L., Scurti, M., Cevenini, E., Castellani, G.C., and Salvioli, S. (2007). Inflammaging and anti-inflammaging: a systemic perspective on aging and longevity emerged from studies in humans. *Mech Ageing Dev* 128**,** 92-105.
167. Francois, R.J., Gardner, D.L., Degrave, E.J., and Bywaters, E.G. (2000). Histopathologic evidence that sacroiliitis in ankylosing spondylitis is not merely enthesitis. *Arthritis Rheum* 43**,** 2011-2024.
168. Frank, D.N., St Amand, A.L., Feldman, R.A., Boedeker, E.C., Harpaz, N., and Pace, N.R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci U S A* 104**,** 13780-13785.
169. Franzosa, E.A., Mciver, L.J., Rahnavard, G., Thompson, L.R., Schirmer, M., Weingart, G., Lipson, K.S., Knight, R., Caporaso, J.G., Segata, N., and Huttenhower, C. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. *Nat Methods* 15**,** 962-968.
170. Franzosa, E.A., Morgan, X.C., Segata, N., Waldron, L., Reyes, J., Earl, A.M., Giannoukos, G., Boylan, M.R., Ciulla, D., Gevers, D., Izard, J., Garrett, W.S., Chan, A.T., and Huttenhower, C. (2014). Relating the metatranscriptome and metagenome of the human gut. *Proc Natl Acad Sci U S A* 111**,** E2329-2338.
171. Fraser, A., Fearon, U., Reece, R., Emery, P., and Veale, D.J. (2001). Matrix metalloproteinase 9, apoptosis, and vascular morphology in early arthritis. *Arthritis Rheum* 44**,** 2024-2028.
172. Frost, G., Sleeth, M.L., Sahuri-Arisoylu, M., Lizarbe, B., Cerdan, S., Brody, L., Anastasovska, J., Ghourab, S., Hankir, M., Zhang, S., Carling, D., Swann, J.R., Gibson, G., Viardot, A., Morrison, D., Louise Thomas, E., and Bell, J.D. (2014). The short-chain fatty acid acetate reduces appetite via a central homeostatic mechanism. *Nat Commun* 5**,** 3611.
173. Fuss, I.J., Heller, F., Boirivant, M., Leon, F., Yoshida, M., Fichtner-Feigl, S., Yang, Z., Exley, M., Kitani, A., Blumberg, R.S., Mannon, P., and Strober, W. (2004). Nonclassical CD1d-restricted NK T cells that produce IL-13 characterize an atypical Th2 response in ulcerative colitis. *J Clin Invest* 113**,** 1490-1497.
174. Gao, X., Lin, H., Revanna, K., and Dong, Q. (2017). A Bayesian taxonomic classification method for 16S rRNA gene sequences with improved species-level accuracy. *BMC Bioinformatics* 18**,** 247.
175. Garn, H., and Renz, H. (2007). Epidemiological and immunological evidence for the hygiene hypothesis. *Immunobiology* 212**,** 441-452.
176. Gevers, D., Kugathasan, S., Denson, L.A., Vazquez-Baeza, Y., Van Treuren, W., Ren, B., Schwager, E., Knights, D., Song, S.J., Yassour, M., Morgan, X.C., Kostic, A.D., Luo, C., Gonzalez, A., Mcdonald, D., Haberman, Y., Walters, T., Baker, S., Rosh, J., Stephens, M., Heyman, M., Markowitz, J., Baldassano, R.,

Griffiths, A., Sylvester, F., Mack, D., Kim, S., Crandall, W., Hyams, J., Huttenhower, C., Knight, R., and Xavier, R.J. (2014). The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* 15**,** 382-392.

177. Gilbert, J.A., Blaser, M.J., Caporaso, J.G., Jansson, J.K., Lynch, S.V., and Knight, R. (2018). Current understanding of the human microbiome. *Nat Med* 24**,** 392-400.
178. Gilbert, J.A., Jansson, J.K., and Knight, R. (2014). The Earth Microbiome project: successes and aspirations. *BMC Biol* 12**,** 69.
179. Gill, T., Brooks, S.R., Rosenbaum, J.T., Asquith, M., and Colbert, R.A. (2019). Novel Inter-omic Analysis Reveals Relationships Between Diverse Gut Microbiota and Host Immune Dysregulation in HLA-B27-Induced Experimental Spondyloarthritis. *Arthritis Rheumatol*.
180. Gilmore, C.P., Donaldson, I., Bo, L., Owens, T., Lowe, J., and Evangelou, N. (2009). Regional variations in the extent and pattern of grey matter demyelination in multiple sclerosis: a comparison between the cerebral cortex, cerebellar cortex, deep grey matter nuclei and the spinal cord. *J Neurol Neurosurg Psychiatry* 80**,** 182-187.
181. Gkouskou, K.K., Deligianni, C., Tsatsanis, C., and Eliopoulos, A.G. (2014). The gut microbiota in mouse models of inflammatory bowel disease. *Front Cell Infect Microbiol* 4**,** 28.
182. Gladman, D.D., Antoni, C., Mease, P., Clegg, D.O., and Nash, P. (2005). Psoriatic arthritis: epidemiology, clinical features, course, and outcome. *Ann Rheum Dis* 64 Suppl 2**,** ii14-17.
183. Gladman, D.D., Shuckett, R., Russell, M.L., Thorne, J.C., and Schachter, R.K. (1987). Psoriatic arthritis (PSA)--an analysis of 220 patients. *Q J Med* 62**,** 127-141.
184. Glintborg, B., Ostergaard, M., Dreyer, L., Krogh, N.S., Tarp, U., Hansen, M.S., Rifbjerg-Madsen, S., Lorenzen, T., and Hetland, M.L. (2011). Treatment response, drug survival, and predictors thereof in 764 patients with psoriatic arthritis treated with anti-tumor necrosis factor alpha therapy: results from the nationwide Danish DANBIO registry. *Arthritis Rheum* 63**,** 382-390.
185. Globig, A.M., Hennecke, N., Martin, B., Seidl, M., Ruf, G., Hasselblatt, P., Thimme, R., and Bengsch, B. (2014). Comprehensive intestinal T helper cell profiling reveals specific accumulation of IFN-gamma+IL-17+coproducing CD4+ T cells in active inflammatory bowel disease. *Inflamm Bowel Dis* 20**,** 2321-2329.
186. Gohl, D.M., Vangay, P., Garbe, J., Maclean, A., Hauge, A., Becker, A., Gould, T.J., Clayton, J.B., Johnson, T.J., Hunter, R., Knights, D., and Beckman, K.B. (2016). Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat Biotechnol* 34**,** 942-949.
187. Goodrich, J.K., Waters, J.L., Poole, A.C., Sutter, J.L., Koren, O., Blekhman, R., Beaumont, M., Van Treuren, W., Knight, R., Bell, J.T., Spector, T.D., Clark, A.G., and Ley, R.E. (2014). Human genetics shape the gut microbiome. *Cell* 159**,** 789-799.
188. Gordon, H.A., and Bruckner-Kardoss, E. (1961). Effect of normal microbial flora on intestinal surface area. *Am J Physiol* 201**,** 175-178.
189. Goto, Y., and Kiyono, H. (2012). Epithelial barrier: an interface for the cross-communication between gut flora and immune system. *Immunol Rev* 245**,** 147-163.

190. Gran, J.T., Husby, G., and Hordvik, M. (1985). Prevalence of ankylosing spondylitis in males and females in a young middle-aged population of Tromso, northern Norway. *Ann Rheum Dis* 44**,** 359-367.
191. Gregersen, P.K., Silver, J., and Winchester, R.J. (1987). The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum* 30**,** 1205-1213.
192. Gregory, S.G., Schmidt, S., Seth, P., Oksenberg, J.R., Hart, J., Prokop, A., Caillier, S.J., Ban, M., Goris, A., Barcellos, L.F., Lincoln, R., Mccauley, J.L., Sawcer, S.J., Compston, D.A., Dubois, B., Hauser, S.L., Garcia-Blanco, M.A., Pericak-Vance, M.A., Haines, J.L., and Multiple Sclerosis Genetics, G. (2007). Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis. *Nat Genet* 39**,** 1083-1091.
193. Gu, S., Fang, L., and Xu, X. (2013). Using SOAPaligner for Short Reads Alignment. *Curr Protoc Bioinformatics* 44**,** 11 11 11-17.
194. Guaraldi, F., and Salvatori, G. (2012). Effect of breast and formula feeding on gut microbiota shaping in newborns. *Front Cell Infect Microbiol* 2**,** 94.
195. Guerrero, R., Margulis, L., and Berlanga, M. (2013). Symbiogenesis: the holobiont as a unit of evolution. *Int Microbiol* 16**,** 133-143.
196. Guglani, L., and Khader, S.A. (2010). Th17 cytokines in mucosal immunity and inflammation. *Curr Opin HIV AIDS* 5**,** 120-127.
197. Guss, J.D., Horsfield, M.W., Fontenele, F.F., Sandoval, T.N., Luna, M., Apoorva, F., Lima, S.F., Bicalho, R.C., Singh, A., Ley, R.E., Van Der Meulen, M.C., Goldring, S.R., and Hernandez, C.J. (2017). Alterations to the Gut Microbiome Impair Bone Strength and Tissue Material Properties. *J Bone Miner Res* 32**,** 1343-1353.
198. Guss, J.D., Taylor, E., Rouse, Z., Roubert, S., Higgins, C.H., Thomas, C.J., Baker, S.P., Vashishth, D., Donnelly, E., Shea, M.K., Booth, S.L., Bicalho, R.C., and Hernandez, C.J. (2019). The microbial metagenome and bone tissue composition in mice with microbiome-induced reductions in bone strength. *Bone* 127**,** 146-154.
199. Haahr, S., Plesner, A.M., Vestergaard, B.F., and Hollsberg, P. (2004). A role of late Epstein-Barr virus infection in multiple sclerosis. *Acta Neurol Scand* 109**,** 270-275.
200. Haahtela, T., Holgate, S., Pawankar, R., Akdis, C.A., Benjaponpitak, S., Caraballo, L., Demain, J., Portnoy, J., Von Hertzen, L., Change, W.a.O.S.C.O.C., and Biodiversity (2013). The biodiversity hypothesis and allergic disease: world allergy organization position statement. *World Allergy Organ J* 6**,** 3.
201. Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S.K., Sodergren, E., Methe, B., Desantis, T.Z., Human Microbiome, C., Petrosino, J.F., Knight, R., and Birren, B.W. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* 21**,** 494-504.
202. Haft, D.H., Loftus, B.J., Richardson, D.L., Yang, F., Eisen, J.A., Paulsen, I.T., and White, O. (2001). TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res* 29**,** 41-43.
203. Haghikia, A., Jorg, S., Duscha, A., Berg, J., Manzel, A., Waschbisch, A., Hammer, A., Lee, D.H., May, C., Wilck, N., Balogh, A., Ostermann, A.I., Schebb, N.H., Akkad, D.A., Grohme, D.A., Kleinewietfeld, M., Kempa, S.,

Thone, J., Demir, S., Muller, D.N., Gold, R., and Linker, R.A. (2015). Dietary Fatty Acids Directly Impact Central Nervous System Autoimmunity via the Small Intestine. *Immunity* 43**,** 817-829.
204. Hagiwara, T., Mukaisho, K., Nakayama, T., Hattori, T., and Sugihara, H. (2015). Proton pump inhibitors and helicobacter pylori-associated pathogenesis. *Asian Pac J Cancer Prev* 16**,** 1315-1319.
205. Haibel, H., Fendler, C., Listing, J., Callhoff, J., Braun, J., and Sieper, J. (2014). Efficacy of oral prednisolone in active ankylosing spondylitis: results of a double-blind, randomised, placebo-controlled short-term trial. *Ann Rheum Dis* 73**,** 243-246.
206. Han, H., Li, Y., Fang, J., Liu, G., Yin, J., Li, T., and Yin, Y. (2018). Gut Microbiota and Type 1 Diabetes. *Int J Mol Sci* 19.
207. Hannon, G.J. (2010). FASTX-Toolkit.
208. Haroon, M., Winchester, R., Giles, J.T., Heffernan, E., and Fitzgerald, O. (2016). Certain class I HLA alleles and haplotypes implicated in susceptibility play a role in determining specific features of the psoriatic arthritis phenotype. *Ann Rheum Dis* 75**,** 155-162.
209. Haroon, N., Inman, R.D., Learch, T.J., Weisman, M.H., Lee, M., Rahbar, M.H., Ward, M.M., Reveille, J.D., and Gensler, L.S. (2013). The impact of tumor necrosis factor alpha inhibitors on radiographic progression in ankylosing spondylitis. *Arthritis Rheum* 65**,** 2645-2654.
210. Heidari, B. (2011). Rheumatoid Arthritis: Early diagnosis and treatment outcomes. *Caspian J Intern Med* 2**,** 161-170.
211. Heller, F., Florian, P., Bojarski, C., Richter, J., Christ, M., Hillenbrand, B., Mankertz, J., Gitter, A.H., Burgel, N., Fromm, M., Zeitz, M., Fuss, I., Strober, W., and Schulzke, J.D. (2005). Interleukin-13 is the key effector Th2 cytokine in ulcerative colitis that affects epithelial tight junctions, apoptosis, and cell restitution. *Gastroenterology* 129**,** 550-564.
212. Helliwell, P.S. (2004). Relationship of psoriatic arthritis with the other spondyloarthropathies. *Curr Opin Rheumatol* 16**,** 344-349.
213. Henderson, P., Anderson, N.H., and Wilson, D.C. (2014). The diagnostic accuracy of fecal calprotectin during the investigation of suspected pediatric inflammatory bowel disease: a systematic review and meta-analysis. *Am J Gastroenterol* 109**,** 637-645.
214. Henke, M.T., Kenny, D.J., Cassilly, C.D., Vlamakis, H., Xavier, R.J., and Clardy, J. (2019). Ruminococcus gnavus, a member of the human gut microbiome associated with Crohn's disease, produces an inflammatory polysaccharide. *Proc Natl Acad Sci U S A* 116**,** 12672-12677.
215. Herfarth, H.H., Kappelman, M.D., Long, M.D., and Isaacs, K.L. (2016). Use of Methotrexate in the Treatment of Inflammatory Bowel Diseases. *Inflamm Bowel Dis* 22**,** 224-233.
216. Herman, S., Kronke, G., and Schett, G. (2008). Molecular mechanisms of inflammatory bone damage: emerging targets for therapy. *Trends Mol Med* 14**,** 245-253.
217. Hernandez-De Sosa, N., Athanasiadis, G., Malouf, J., Laiz, A., Marin, A., Herrera, S., Farrerons, J., Soria, J.M., and Casademont, J. (2014). Heritability of bone mineral density in a multivariate family-based study. *Calcif Tissue Int* 94**,** 590-596.

218. Hernlund, E., Svedbom, A., Ivergard, M., Compston, J., Cooper, C., Stenmark, J., Mccloskey, E.V., Jonsson, B., and Kanis, J.A. (2013). Osteoporosis in the European Union: medical management, epidemiology and economic burden. A report prepared in collaboration with the International Osteoporosis Foundation (IOF) and the European Federation of Pharmaceutical Industry Associations (EFPIA). *Arch Osteoporos* 8**,** 136.
219. Hesper, B., and Hogeweg, P. (1970). Bioinformatica: een werkconcept. *Kameleon* 1.
220. Highton, J., Stebbings, S., and Jenks, K. (2012). The intestinal microbiome in ankylosing spondylitis: comment on the article by Rosenbaum and Davey. *Arthritis Rheum* 64**,** 597-598; author reply 598.
221. Hillmann, B., Al-Ghalith, G.A., Shields-Cutler, R.R., Zhu, Q., Gohl, D.M., Beckman, K.B., Knight, R., and Knights, D. (2018). Evaluating the Information Content of Shallow Shotgun Metagenomics. *mSystems* 3.
222. Hofbauer, L.C., and Schoppet, M. (2004). Clinical implications of the osteoprotegerin/RANKL/RANK system for bone and vascular diseases. *JAMA* 292**,** 490-495.
223. Hosseini, E., Grootaert, C., Verstraete, W., and Van De Wiele, T. (2011). Propionate as a health-promoting microbial metabolite in the human gut. *Nutr Rev* 69**,** 245-258.
224. Hotamisligil, G.S. (2017). Inflammation, metaflammation and immunometabolic disorders. *Nature* 542**,** 177-185.
225. Hou, J.K., Abraham, B., and El-Serag, H. (2011). Dietary intake and risk of developing inflammatory bowel disease: a systematic review of the literature. *Am J Gastroenterol* 106**,** 563-573.
226. Houssien, D.A., Jonsson, T., Davies, E., and Scott, D.L. (1997). Clinical significance of IgA rheumatoid factor subclasses in rheumatoid arthritis. *J Rheumatol* 24**,** 2119-2122.
227. Hsu, E., and Pacifici, R. (2018). From Osteoimmunology to Osteomicrobiology: How the Microbiota and the Immune System Regulate Bone. *Calcif Tissue Int* 102**,** 512-521.
228. Huffmeier, U., Uebe, S., Ekici, A.B., Bowes, J., Giardina, E., Korendowych, E., Juneblad, K., Apel, M., Mcmanus, R., Ho, P., Bruce, I.N., Ryan, A.W., Behrens, F., Lascorz, J., Bohm, B., Traupe, H., Lohmann, J., Gieger, C., Wichmann, H.E., Herold, C., Steffens, M., Klareskog, L., Wienker, T.F., Fitzgerald, O., Alenius, G.M., Mchugh, N.J., Novelli, G., Burkhardt, H., Barton, A., and Reis, A. (2010). Common variants at TRAF3IP2 are associated with susceptibility to psoriatic arthritis and psoriasis. *Nat Genet* 42**,** 996-999.
229. Hughes, J.B., and Hellmann, J.J. (2005). The application of rarefaction techniques to molecular inventories of microbial diversity. *Methods Enzymol* 397**,** 292-308.
230. Hugot, J.P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J.P., Belaiche, J., Almer, S., Tysk, C., O'morain, C.A., Gassull, M., Binder, V., Finkel, Y., Cortot, A., Modigliani, R., Laurent-Puig, P., Gower-Rousseau, C., Macry, J., Colombel, J.F., Sahbatou, M., and Thomas, G. (2001). Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 411**,** 599-603.
231. Hukuda, S., Minami, M., Saito, T., Mitsui, H., Matsui, N., Komatsubara, Y., Makino, H., Shibata, T., Shingu, M., Sakou, T., and Shichikawa, K. (2001).

Spondyloarthropathies in Japan: nationwide questionnaire survey performed by the Japan Ankylosing Spondylitis Society. *J Rheumatol* 28**,** 554-559.
232. Human Microbiome Project, C. (2012). Structure, function and diversity of the healthy human microbiome. *Nature* 486**,** 207-214.
233. Huson, D.H., and Xie, C. (2014). A poor man's BLASTX--high-throughput metagenomic protein database search using PAUDA. *Bioinformatics* 30**,** 38-39.
234. Hyatt, D., Locascio, P.F., Hauser, L.J., and Uberbacher, E.C. (2012). Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* 28**,** 2223-2230.
235. Imam, T., Park, S., Kaplan, M.H., and Olson, M.R. (2018). Effector T Helper Cell Subsets in Inflammatory Bowel Diseases. *Front Immunol* 9**,** 1212.
236. Inohara, N., Ogura, Y., Fontalba, A., Gutierrez, O., Pons, F., Crespo, J., Fukase, K., Inamura, S., Kusumoto, S., Hashimoto, M., Foster, S.J., Moran, A.P., Fernandez-Luna, J.L., and Nunez, G. (2003). Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease. *J Biol Chem* 278**,** 5509-5512.
237. Inoue, N., Tamura, K., Kinouchi, Y., Fukuda, Y., Takahashi, S., Ogura, Y., Inohara, N., Nunez, G., Kishi, Y., Koike, Y., Shimosegawa, T., Shimoyama, T., and Hibi, T. (2002). Lack of common NOD2 variants in Japanese patients with Crohn's disease. *Gastroenterology* 123**,** 86-91.
238. Integrative, H.M.P.R.N.C. (2014). The Integrative Human Microbiome Project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease. *Cell Host Microbe* 16**,** 276-289.
239. International Genetics of Ankylosing Spondylitis, C., Cortes, A., Hadler, J., Pointon, J.P., Robinson, P.C., Karaderi, T., Leo, P., Cremin, K., Pryce, K., Harris, J., Lee, S., Joo, K.B., Shim, S.C., Weisman, M., Ward, M., Zhou, X., Garchon, H.J., Chiocchia, G., Nossent, J., Lie, B.A., Forre, O., Tuomilehto, J., Laiho, K., Jiang, L., Liu, Y., Wu, X., Bradbury, L.A., Elewaut, D., Burgos-Vargas, R., Stebbings, S., Appleton, L., Farrah, C., Lau, J., Kenna, T.J., Haroon, N., Ferreira, M.A., Yang, J., Mulero, J., Fernandez-Sueiro, J.L., Gonzalez-Gay, M.A., Lopez-Larrea, C., Deloukas, P., Donnelly, P., Australo-Anglo-American Spondyloarthritis, C., Groupe Francaise D'etude Genetique Des, S., Nord-Trondelag Health, S., Spondyloarthritis Research Consortium Of, C., Wellcome Trust Case Control, C., Bowness, P., Gafney, K., Gaston, H., Gladman, D.D., Rahman, P., Maksymowych, W.P., Xu, H., Crusius, J.B., Van Der Horst-Bruinsma, I.E., Chou, C.T., Valle-Onate, R., Romero-Sanchez, C., Hansen, I.M., Pimentel-Santos, F.M., Inman, R.D., Videm, V., Martin, J., Breban, M., Reveille, J.D., Evans, D.M., Kim, T.H., Wordsworth, B.P., and Brown, M.A. (2013). Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat Genet* 45**,** 730-738.
240. International Multiple Sclerosis Genetics, C., Hafler, D.A., Compston, A., Sawcer, S., Lander, E.S., Daly, M.J., De Jager, P.L., De Bakker, P.I., Gabriel, S.B., Mirel, D.B., Ivinson, A.J., Pericak-Vance, M.A., Gregory, S.G., Rioux, J.D., Mccauley, J.L., Haines, J.L., Barcellos, L.F., Cree, B., Oksenberg, J.R., and Hauser, S.L. (2007). Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med* 357**,** 851-862.

241. Iseme, R.A., Mcevoy, M., Kelly, B., Agnew, L., Walker, F.R., and Attia, J. (2017). Is osteoporosis an autoimmune mediated disorder? *Bone Rep* 7**,** 121-131.
242. Ivady, G., Madar, L., Dzsudzsak, E., Koczok, K., Kappelmayer, J., Krulisova, V., Macek, M., Jr., Horvath, A., and Balogh, I. (2018). Analytical parameters and validation of homopolymer detection in a pyrosequencing-based next generation sequencing system. *BMC Genomics* 19**,** 158.
243. Iwai, S., Weinmaier, T., Schmidt, B.L., Albertson, D.G., Poloso, N.J., Dabbagh, K., and Desantis, T.Z. (2016). Piphillin: Improved Prediction of Metagenomic Content by Direct Inference from Human Microbiomes. *PLoS One* 11**,** e0166104.
244. Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin del la Société Vaudoise des Sciences Naturelles* 37**,** 547--579.
245. Jackson, M.A., Goodrich, J.K., Maxan, M.E., Freedberg, D.E., Abrams, J.A., Poole, A.C., Sutter, J.L., Welter, D., Ley, R.E., Bell, J.T., Spector, T.D., and Steves, C.J. (2016). Proton pump inhibitors alter the composition of the gut microbiota. *Gut* 65**,** 749-756.
246. Jackson, M.A., Verdi, S., Maxan, M.E., Shin, C.M., Zierer, J., Bowyer, R.C.E., Martin, T., Williams, F.M.K., Menni, C., Bell, J.T., Spector, T.D., and Steves, C.J. (2018). Gut microbiota associations with common diseases and prescription medications in a population-based cohort. *Nat Commun* 9**,** 2655.
247. Janeiro, M.H., Ramirez, M.J., Milagro, F.I., Martinez, J.A., and Solas, M. (2018). Implication of Trimethylamine N-Oxide (TMAO) in Disease: Potential Biomarker or New Therapeutic Target. *Nutrients* 10.
248. Jangi, S., Gandhi, R., Cox, L.M., Li, N., Von Glehn, F., Yan, R., Patel, B., Mazzola, M.A., Liu, S., Glanz, B.L., Cook, S., Tankou, S., Stuart, F., Melo, K., Nejad, P., Smith, K., Topcuolu, B.D., Holden, J., Kivisakk, P., Chitnis, T., De Jager, P.L., Quintana, F.J., Gerber, G.K., Bry, L., and Weiner, H.L. (2016). Alterations of the human gut microbiome in multiple sclerosis. *Nat Commun* 7**,** 12015.
249. Jeong, Y., Kim, J.W., You, H.J., Park, S.J., Lee, J., Ju, J.H., Park, M.S., Jin, H., Cho, M.L., Kwon, B., Park, S.H., and Ji, G.E. (2019). Gut Microbial Composition and Function Are Altered in Patients with Early Rheumatoid Arthritis. *J Clin Med* 8.
250. Jernberg, C., Lofmark, S., Edlund, C., and Jansson, J.K. (2010). Long-term impacts of antibiotic exposure on the human intestinal microbiota. *Microbiology* 156**,** 3216-3223.
251. Jethwa, H., and Bowness, P. (2016). The interleukin (IL)-23/IL-17 axis in ankylosing spondylitis: new advances and potentials for treatment. *Clin Exp Immunol* 183**,** 30-36.
252. Jia, W., Xie, G., and Jia, W. (2018). Bile acid-microbiota crosstalk in gastrointestinal inflammation and carcinogenesis. *Nat Rev Gastroenterol Hepatol* 15**,** 111-128.
253. Jie, Z., Xia, H., Zhong, S.L., Feng, Q., Li, S., Liang, S., Zhong, H., Liu, Z., Gao, Y., Zhao, H., Zhang, D., Su, Z., Fang, Z., Lan, Z., Li, J., Xiao, L., Li, J., Li, R., Li, X., Li, F., Ren, H., Huang, Y., Peng, Y., Li, G., Wen, B., Dong, B., Chen, J.Y., Geng, Q.S., Zhang, Z.W., Yang, H., Wang, J., Wang, J., Zhang, X., Madsen, L., Brix, S., Ning, G., Xu, X., Liu, X., Hou, Y., Jia, H., He, K., and

Kristiansen, K. (2017). The gut microbiome in atherosclerotic cardiovascular disease. *Nat Commun* 8**,** 845.
254. Jin, D., Wu, S., Zhang, Y.G., Lu, R., Xia, Y., Dong, H., and Sun, J. (2015). Lack of Vitamin D Receptor Causes Dysbiosis and Changes the Functions of the Murine Intestinal Microbiome. *Clin Ther* 37**,** 996-1009 e1007.
255. Johnell, O., and Kanis, J.A. (2006). An estimate of the worldwide prevalence and disability associated with osteoporotic fractures. *Osteoporos Int* 17**,** 1726-1733.
256. Johnson, A.J., Vangay, P., Al-Ghalith, G.A., Hillmann, B.M., Ward, T.L., Shields-Cutler, R.R., Kim, A.D., Shmagel, A.K., Syed, A.N., Personalized Microbiome Class, S., Walter, J., Menon, R., Koecher, K., and Knights, D. (2019). Daily Sampling Reveals Personalized Diet-Microbiome Associations in Humans. *Cell Host Microbe* 25**,** 789-802 e785.
257. Jones, R.M., Mulle, J.G., and Pacifici, R. (2018). Osteomicrobiology: The influence of gut microbiota on bone in health and disease. *Bone* 115**,** 59-67.
258. Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., Mcgovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A., Essers, J., Mitrovic, M., Ning, K., Cleynen, I., Theatre, E., Spain, S.L., Raychaudhuri, S., Goyette, P., Wei, Z., Abraham, C., Achkar, J.P., Ahmad, T., Amininejad, L., Ananthakrishnan, A.N., Andersen, V., Andrews, J.M., Baidoo, L., Balschun, T., Bampton, P.A., Bitton, A., Boucher, G., Brand, S., Buning, C., Cohain, A., Cichon, S., D'amato, M., De Jong, D., Devaney, K.L., Dubinsky, M., Edwards, C., Ellinghaus, D., Ferguson, L.R., Franchimont, D., Fransen, K., Gearry, R., Georges, M., Gieger, C., Glas, J., Haritunians, T., Hart, A., Hawkey, C., Hedl, M., Hu, X., Karlsen, T.H., Kupcinskas, L., Kugathasan, S., Latiano, A., Laukens, D., Lawrance, I.C., Lees, C.W., Louis, E., Mahy, G., Mansfield, J., Morgan, A.R., Mowat, C., Newman, W., Palmieri, O., Ponsioen, C.Y., Potocnik, U., Prescott, N.J., Regueiro, M., Rotter, J.I., Russell, R.K., Sanderson, J.D., Sans, M., Satsangi, J., Schreiber, S., Simms, L.A., Sventoraityte, J., Targan, S.R., Taylor, K.D., Tremelling, M., Verspaget, H.W., De Vos, M., Wijmenga, C., Wilson, D.C., Winkelmann, J., Xavier, R.J., Zeissig, S., Zhang, B., Zhang, C.K., Zhao, H., International, I.B.D.G.C., Silverberg, M.S., Annese, V., Hakonarson, H., Brant, S.R., Radford-Smith, G., Mathew, C.G., Rioux, J.D., et al. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 491**,** 119-124.
259. Jovel, J., Patterson, J., Wang, W., Hotte, N., O'keefe, S., Mitchel, T., Perry, T., Kao, D., Mason, A.L., Madsen, K.L., and Wong, G.K. (2016). Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics. *Front Microbiol* 7**,** 459.
260. Kaipiainen-Seppanen, O., Aho, K., and Heliovaara, M. (1997). Incidence and prevalence of ankylosing spondylitis in Finland. *J Rheumatol* 24**,** 496-499.
261. Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28**,** 27-30.
262. Kang, D.D., Froula, J., Egan, R., and Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3**,** e1165.
263. Kang, S., Denman, S.E., Morrison, M., Yu, Z., Dore, J., Leclerc, M., and Mcsweeney, C.S. (2010). Dysbiosis of fecal microbiota in Crohn's disease

patients as revealed by a custom phylogenetic microarray. *Inflamm Bowel Dis* 16**,** 2034-2042.
264. Kanis, J.A., Johnell, O., Oden, A., Johansson, H., and Mccloskey, E. (2008). FRAX and the assessment of fracture probability in men and women from the UK. *Osteoporos Int* 19**,** 385-397.
265. Kanis, J.A., Mccloskey, E.V., Johansson, H., Cooper, C., Rizzoli, R., Reginster, J.Y., Scientific Advisory Board of the European Society For, C., Economic Aspects Of, O., Osteoarthritis, and The Committee of Scientific Advisors of the International Osteoporosis, F. (2013). European guidance for the diagnosis and management of osteoporosis in postmenopausal women. *Osteoporos Int* 24**,** 23-57.
266. Kanis, J.A., Melton, L.J., 3rd, Christiansen, C., Johnston, C.C., and Khaltaev, N. (1994). The diagnosis of osteoporosis. *J Bone Miner Res* 9**,** 1137-1141.
267. Karstens, L., Asquith, M., Davin, S., Fair, D., Gregory, W.T., Wolfe, A.J., Braun, J., and Mcweeney, S. (2019). Controlling for Contaminants in Low-Biomass 16S rRNA Gene Sequencing Experiments. *mSystems* 4.
268. Kaser, A., and Blumberg, R.S. (2011). Autophagy, microbial sensing, endoplasmic reticulum stress, and epithelial function in inflammatory bowel disease. *Gastroenterology* 140**,** 1738-1747.
269. Kato, K., Mizuno, S., Umesaki, Y., Ishii, Y., Sugitani, M., Imaoka, A., Otsuka, M., Hasunuma, O., Kurihara, R., Iwasaki, A., and Arakawa, Y. (2004). Randomized placebo-controlled trial assessing the effect of bifidobacteria-fermented milk on active ulcerative colitis. *Aliment Pharmacol Ther* 20**,** 1133-1141.
270. Kembel, S.W., Wu, M., Eisen, J.A., and Green, J.L. (2012). Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Comput Biol* 8**,** e1002743.
271. Kennedy, E.A., King, K.Y., and Baldridge, M.T. (2018). Mouse Microbiota Models: Comparing Germ-Free Mice and Antibiotics Treatment as Tools for Modifying Gut Bacteria. *Front Physiol* 9**,** 1534.
272. Keyszer, G., Redlich, A., Haupl, T., Zacher, J., Sparmann, M., Engethum, U., Gay, S., and Burmester, G.R. (1998). Differential expression of cathepsins B and L compared with matrix metalloproteinases and their respective inhibitors in rheumatoid arthritis and osteoarthritis: a parallel investigation by semiquantitative reverse transcriptase-polymerase chain reaction and immunohistochemistry. *Arthritis Rheum* 41**,** 1378-1387.
273. Kho, Z.Y., and Lal, S.K. (2018). The Human Gut Microbiome - A Potential Controller of Wellness and Disease. *Front Microbiol* 9**,** 1835.
274. Khosla, S., Bilezikian, J.P., Dempster, D.W., Lewiecki, E.M., Miller, P.D., Neer, R.M., Recker, R.R., Shane, E., Shoback, D., and Potts, J.T. (2012). Benefits and risks of bisphosphonate therapy for osteoporosis. *J Clin Endocrinol Metab* 97**,** 2272-2282.
275. Khundmiri, S.J., Murray, R.D., and Lederer, E. (2016). PTH and Vitamin D. *Compr Physiol* 6**,** 561-601.
276. Kim, K.S., Hong, S.W., Han, D., Yi, J., Jung, J., Yang, B.G., Lee, J.Y., Lee, M., and Surh, C.D. (2016). Dietary antigens limit mucosal immunity by inducing regulatory T cells in the small intestine. *Science* 351**,** 858-863.
277. Kingwell, E., Marriott, J.J., Jette, N., Pringsheim, T., Makhani, N., Morrow, S.A., Fisk, J.D., Evans, C., Beland, S.G., Kulaga, S., Dykeman, J., Wolfson, C.,

Koch, M.W., and Marrie, R.A. (2013). Incidence and prevalence of multiple sclerosis in Europe: a systematic review. *BMC Neurol* 13**,** 128.
278. Kinne, R.W., Brauer, R., Stuhlmuller, B., Palombo-Kinne, E., and Burmester, G.R. (2000). Macrophages in rheumatoid arthritis. *Arthritis Res* 2**,** 189-202.
279. Kitz, A., Singer, E., and Hafler, D. (2018). Regulatory T Cells: From Discovery to Autoimmunity. *Cold Spring Harb Perspect Med* 8.
280. Klarenbeek, P.L., De Hair, M.J., Doorenspleet, M.E., Van Schaik, B.D., Esveldt, R.E., Van De Sande, M.G., Cantaert, T., Gerlag, D.M., Baeten, D., Van Kampen, A.H., Baas, F., Tak, P.P., and De Vries, N. (2012). Inflamed target tissue provides a specific niche for highly expanded T-cell clones in early human autoimmune disease. *Ann Rheum Dis* 71**,** 1088-1093.
281. Klingenberg, H., Asshauer, K.P., Lingner, T., and Meinicke, P. (2013). Protein signature-based estimation of metagenomic abundances including all domains of life and viruses. *Bioinformatics* 29**,** 973-980.
282. Knight, R., Callewaert, C., Marotz, C., Hyde, E.R., Debelius, J.W., Mcdonald, D., and Sogin, M.L. (2017). The Microbiome and Human Biology. *Annu Rev Genomics Hum Genet* 18**,** 65-86.
283. Kobak, S., and Bes, C. (2018). An autumn tale: geriatric rheumatoid arthritis. *Ther Adv Musculoskelet Dis* 10**,** 3-11.
284. Kobayashi, T., Okamoto, S., Hisamatsu, T., Kamada, N., Chinen, H., Saito, R., Kitazume, M.T., Nakazawa, A., Sugita, A., Koganei, K., Isobe, K., and Hibi, T. (2008). IL23 differentially regulates the Th1/Th17 balance in ulcerative colitis and Crohn's disease. *Gut* 57**,** 1682-1689.
285. Kopylova, E., Noe, L., and Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28**,** 3211-3217.
286. Kraemer, W.J., Hatfield, D.L., Comstock, B.A., Fragala, M.S., Davitt, P.M., Cortis, C., Wilson, J.M., Lee, E.C., Newton, R.U., Dunn-Lewis, C., Hakkinen, K., Szivak, T.K., Hooper, D.R., Flanagan, S.D., Looney, D.P., White, M.T., Volek, J.S., and Maresh, C.M. (2014). Influence of HMB supplementation and resistance training on cytokine responses to resistance exercise. *J Am Coll Nutr* 33**,** 247-255.
287. Kruithof, E., Baeten, D., De Rycke, L., Vandooren, B., Foell, D., Roth, J., Canete, J.D., Boots, A.M., Veys, E.M., and De Keyser, F. (2005). Synovial histopathology of psoriatic arthritis, both oligo- and polyarticular, resembles spondyloarthropathy more than it does rheumatoid arthritis. *Arthritis Res Ther* 7**,** R569-580.
288. Kultima, J.R., Coelho, L.P., Forslund, K., Huerta-Cepas, J., Li, S.S., Driessen, M., Voigt, A.Y., Zeller, G., Sunagawa, S., and Bork, P. (2016). MOCAT2: a metagenomic assembly, annotation and profiling framework. *Bioinformatics* 32**,** 2520-2523.
289. Kurowska-Stolarska, M., and Alivernini, S. (2017). Synovial tissue macrophages: friend or foe? *RMD Open* 3**,** e000527.
290. Kurtzke, J.F. (1983). Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 33**,** 1444-1452.
291. Lakatos, P.L., Szamosi, T., and Lakatos, L. (2007). Smoking in inflammatory bowel diseases: good, bad or ugly? *World J Gastroenterol* 13**,** 6134-6139.
292. Langdahl, B.L., Uitterlinden, A.G., Ralston, S.H., Trikalinos, T.A., Balcells, S., Brandi, M.L., Scollen, S., Lips, P., Lorenc, R., Obermayer-Pietsch, B., Reid,

D.M., Armas, J.B., Arp, P.P., Bassiti, A., Bustamante, M., Husted, L.B., Carey, A.H., Perez Cano, R., Dobnig, H., Dunning, A.M., Fahrleitner-Pammer, A., Falchetti, A., Karczmarewicz, E., Kruk, M., Van Leeuwen, J.P., Masi, L., Van Meurs, J.B., Mangion, J., Mcguigan, F.E., Mellibovsky, L., Mosekilde, L., Nogues, X., Pols, H.A., Reeve, J., Renner, W., Rivadeneira, F., Van Schoor, N.M., Ioannidis, J.P., Investigators, A., Investigators, D., Investigators, E., Investigators, E., Investigators, F., Investigators, L., Investigators, E., and Study, G. (2008). Large-scale analysis of association between polymorphisms in the transforming growth factor beta 1 gene (TGFB1) and osteoporosis: the GENOMOS study. *Bone* 42**,** 969-981.

293. Langille, M.G., Zaneveld, J., Caporaso, J.G., Mcdonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Vega Thurber, R.L., Knight, R., Beiko, R.G., and Huttenhower, C. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* 31**,** 814-821.
294. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9**,** 357-359.
295. Lawley, T.D., Clare, S., Walker, A.W., Stares, M.D., Connor, T.R., Raisen, C., Goulding, D., Rad, R., Schreiber, F., Brandt, C., Deakin, L.J., Pickard, D.J., Duncan, S.H., Flint, H.J., Clark, T.G., Parkhill, J., and Dougan, G. (2012). Targeted restoration of the intestinal microbiota with a simple, defined bacteriotherapy resolves relapsing Clostridium difficile disease in mice. *PLoS Pathog* 8**,** e1002995.
296. Le Roy, C.I., Beaumont, M., Jackson, M.A., Steves, C.J., Spector, T.D., and Bell, J.T. (2018). Heritable components of the human fecal microbiome are associated with visceral fat. *Gut Microbes* 9**,** 61-67.
297. Leblanc, J.G., Milani, C., De Giori, G.S., Sesma, F., Van Sinderen, D., and Ventura, M. (2013). Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Curr Opin Biotechnol* 24**,** 160-168.
298. Lebwohl, B., Cao, Y., Zong, G., Hu, F.B., Green, P.H.R., Neugut, A.I., Rimm, E.B., Sampson, L., Dougherty, L.W., Giovannucci, E., Willett, W.C., Sun, Q., and Chan, A.T. (2017). Long term gluten consumption in adults without celiac disease and risk of coronary heart disease: prospective cohort study. *BMJ* 357**,** j1892.
299. Lecerf, J.M., Depeint, F., Clerc, E., Dugenet, Y., Niamba, C.N., Rhazi, L., Cayzeele, A., Abdelnour, G., Jaruga, A., Younes, H., Jacobs, H., Lambrey, G., Abdelnour, A.M., and Pouillart, P.R. (2012). Xylo-oligosaccharide (XOS) in combination with inulin modulates both the intestinal environment and immune status in healthy subjects, while XOS alone only shows prebiotic properties. *Br J Nutr* 108**,** 1847-1858.
300. Lederberg, J., and Mccray, A.T. (2001). Ome Sweet 'Omics-- A Genealogical Treasury of Words *The Scientist***,** p7.
301. Lee, D., Albenberg, L., Compher, C., Baldassano, R., Piccoli, D., Lewis, J.D., and Wu, G.D. (2015). Diet in the pathogenesis and treatment of inflammatory bowel diseases. *Gastroenterology* 148**,** 1087-1106.
302. Lee, N., and Kim, W.U. (2017). Microbiota in T-cell homeostasis and inflammatory diseases. *Exp Mol Med* 49**,** e340.
303. Lee, Y.K., Menezes, J.S., Umesaki, Y., and Mazmanian, S.K. (2011). Proinflammatory T-cell responses to gut microbiota promote experimental

autoimmune encephalomyelitis. *Proc Natl Acad Sci U S A* 108 Suppl 1**,** 4615-4622.
304. Lefevre, S., Meier, F.M., Neumann, E., and Muller-Ladner, U. (2015). Role of synovial fibroblasts in rheumatoid arthritis. *Curr Pharm Des* 21**,** 130-141.
305. Leong, R.W., Armuzzi, A., Ahmad, T., Wong, M.L., Tse, P., Jewell, D.P., and Sung, J.J. (2003). NOD2/CARD15 gene polymorphisms and Crohn's disease in the Chinese population. *Aliment Pharmacol Ther* 17**,** 1465-1470.
306. Leslie, W.D., Miller, N., Rogala, L., and Bernstein, C.N. (2008). Vitamin D status and bone density in recently diagnosed inflammatory bowel disease: the Manitoba IBD Cohort Study. *Am J Gastroenterol* 103**,** 1451-1459.
307. Levene, M.J., Korlach, J., Turner, S.W., Foquet, M., Craighead, H.G., and Webb, W.W. (2003). Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* 299**,** 682-686.
308. Lewiecki, E.M. (2010). Bisphosphonates for the treatment of osteoporosis: insights for clinicians. *Ther Adv Chronic Dis* 1**,** 115-128.
309. Li, C., Huang, Q., Yang, R., Dai, Y., Zeng, Y., Tao, L., Li, X., Zeng, J., and Wang, Q. (2019). Gut microbiota composition and bone mineral loss-epidemiologic evidence from individuals in Wuhan, China. *Osteoporos Int* 30**,** 1003-1013.
310. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25**,** 1754-1760.
311. Li, J., Ueno, A., Fort Gasia, M., Luider, J., Wang, T., Hirota, C., Jijon, H.B., Deane, M., Tom, M., Chan, R., Barkema, H.W., Beck, P.L., Kaplan, G.G., Panaccione, R., Qian, J., Iacucci, M., Gui, X., and Ghosh, S. (2016). Profiles of Lamina Propria T Helper Cell Subsets Discriminate Between Ulcerative Colitis and Crohn's Disease. *Inflamm Bowel Dis* 22**,** 1779-1792.
312. Li, R., Patterson, K.R., and Bar-Or, A. (2018). Reassessing B cell contributions in multiple sclerosis. *Nat Immunol* 19**,** 696-707.
313. Liao, Z.T., Pan, Y.F., Huang, J.L., Huang, F., Chi, W.J., Zhang, K.X., Lin, Z.M., Wu, Y.Q., He, W.Z., Wu, J., Xie, X.J., Huang, J.X., Wei, Q.J., Li, T.W., Wu, Z., Yu, B.Y., and Gu, J.R. (2009). An epidemiological survey of low back pain and axial spondyloarthritis in a Chinese Han population. *Scand J Rheumatol* 38**,** 455-459.
314. Lin, P., Bach, M., Asquith, M., Lee, A.Y., Akileswaran, L., Stauffer, P., Davin, S., Pan, Y., Cambronne, E.D., Dorris, M., Debelius, J.W., Lauber, C.L., Ackermann, G., Baeza, Y.V., Gill, T., Knight, R., Colbert, R.A., Taurog, J.D., Van Gelder, R.N., and Rosenbaum, J.T. (2014). HLA-B27 and human beta2-microglobulin affect the gut microbiota of transgenic rats. *PLoS One* 9**,** e105684.
315. Liu, J.T., Yeh, H.M., Liu, S.Y., and Chen, K.T. (2014). Psoriatic arthritis: Epidemiology, diagnosis, and treatment. *World J Orthop* 5**,** 537-543.
316. Liu, X., Zeng, B., Zhang, J., Li, W., Mou, F., Wang, H., Zou, Q., Zhong, B., Wu, L., Wei, H., and Fang, Y. (2016). Role of the Gut Microbiome in Modulating Arthritis Progression in Mice. *Sci Rep* 6**,** 30594.
317. Liu, Y., Alookaran, J.J., and Rhoads, J.M. (2018). Probiotics in Autoimmune and Inflammatory Disorders. *Nutrients* 10.
318. Lloyd-Price, J., Arze, C., Ananthakrishnan, A.N., Schirmer, M., Avila-Pacheco, J., Poon, T.W., Andrews, E., Ajami, N.J., Bonham, K.S., Brislawn, C.J., Casero, D., Courtney, H., Gonzalez, A., Graeber, T.G., Hall, A.B., Lake, K., Landers, C.J., Mallick, H., Plichta, D.R., Prasad, M., Rahnavard, G., Sauk, J., Shungin,

D., Vazquez-Baeza, Y., White, R.A., 3rd, Investigators, I., Braun, J., Denson, L.A., Jansson, J.K., Knight, R., Kugathasan, S., Mcgovern, D.P.B., Petrosino, J.F., Stappenbeck, T.S., Winter, H.S., Clish, C.B., Franzosa, E.A., Vlamakis, H., Xavier, R.J., and Huttenhower, C. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* 569**,** 655-662.

319. Loddo, I., and Romano, C. (2015). Inflammatory Bowel Disease: Genetics, Epigenetics, and Pathogenesis. *Front Immunol* 6**,** 551.
320. Louis, E., Irving, P., and Beaugerie, L. (2014). Use of azathioprine in IBD: modern aspects of an old drug. *Gut* 63**,** 1695-1699.
321. Louis, P., and Flint, H.J. (2017). Formation of propionate and butyrate by the human colonic microbiota. *Environ Microbiol* 19**,** 29-41.
322. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15**,** 550.
323. Lozupone, C., and Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 71**,** 8228-8235.
324. Lublin, F.D., Reingold, S.C., Cohen, J.A., Cutter, G.R., Sorensen, P.S., Thompson, A.J., Wolinsky, J.S., Balcer, L.J., Banwell, B., Barkhof, F., Bebo, B., Jr., Calabresi, P.A., Clanet, M., Comi, G., Fox, R.J., Freedman, M.S., Goodman, A.D., Inglese, M., Kappos, L., Kieseier, B.C., Lincoln, J.A., Lubetzki, C., Miller, A.E., Montalban, X., O'connor, P.W., Petkau, J., Pozzilli, C., Rudick, R.A., Sormani, M.P., Stuve, O., Waubant, E., and Polman, C.H. (2014). Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology* 83**,** 278-286.
325. Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S.M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T.W., and Wang, J. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1**,** 18.
326. Luthold, R.V., Fernandes, G.R., Franco-De-Moraes, A.C., Folchetti, L.G., and Ferreira, S.R. (2017). Gut microbiota interactions with the immunomodulatory role of vitamin D in normal individuals. *Metabolism* 69**,** 76-86.
327. Macgregor, A.J., Snieder, H., Rigby, A.S., Koskenvuo, M., Kaprio, J., Aho, K., and Silman, A.J. (2000). Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum* 43**,** 30-37.
328. Machiels, K., Joossens, M., Sabino, J., De Preter, V., Arijs, I., Eeckhaut, V., Ballet, V., Claes, K., Van Immerseel, F., Verbeke, K., Ferrante, M., Verhaegen, J., Rutgeerts, P., and Vermeire, S. (2014). A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. *Gut* 63**,** 1275-1283.
329. Mackie, R.I., Sghir, A., and Gaskins, H.R. (1999). Developmental microbial ecology of the neonatal gastrointestinal tract. *Am J Clin Nutr* 69**,** 1035S-1045S.
330. Magoc, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27**,** 2957-2963.
331. Magyari, L., Varszegi, D., Kovesdi, E., Sarlos, P., Farago, B., Javorhazy, A., Sumegi, K., Banfai, Z., and Melegh, B. (2014). Interleukins and interleukin receptors in rheumatoid arthritis: Research, diagnostics and clinical implications. *World J Orthop* 5**,** 516-536.

332. Maier, L., Pruteanu, M., Kuhn, M., Zeller, G., Telzerow, A., Anderson, E.E., Brochado, A.R., Fernandez, K.C., Dose, H., Mori, H., Patil, K.R., Bork, P., and Typas, A. (2018). Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature* 555**,** 623-628.
333. Makrygiannakis, D., Hermansson, M., Ulfgren, A.K., Nicholas, A.P., Zendman, A.J., Eklund, A., Grunewald, J., Skold, C.M., Klareskog, L., and Catrina, A.I. (2008). Smoking increases peptidylarginine deiminase 2 enzyme expression in human lungs and increases citrullination in BAL cells. *Ann Rheum Dis* 67**,** 1488-1492.
334. Malmstrom, V., Catrina, A.I., and Klareskog, L. (2017). The immunopathogenesis of seropositive rheumatoid arthritis: from triggering to targeting. *Nat Rev Immunol* 17**,** 60-75.
335. Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., Nalin, R., Jarrin, C., Chardon, P., Marteau, P., Roca, J., and Dore, J. (2006). Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* 55**,** 205-211.
336. Marks, D.J., Harbord, M.W., Macallister, R., Rahman, F.Z., Young, J., Al-Lazikani, B., Lees, W., Novelli, M., Bloom, S., and Segal, A.W. (2006). Defective acute inflammation in Crohn's disease: a clinical investigation. *Lancet* 367**,** 668-678.
337. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* 17**,** 3.
338. Martinez, I., Stegen, J.C., Maldonado-Gomez, M.X., Eren, A.M., Siba, P.M., Greenhill, A.R., and Walter, J. (2015). The gut microbiota of rural papua new guineans: composition, diversity patterns, and ecological processes. *Cell Rep* 11**,** 527-538.
339. Masella, A.P., Bartram, A.K., Truszkowski, J.M., Brown, D.G., and Neufeld, J.D. (2012). PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* 13**,** 31.
340. Maslowski, K.M., Vieira, A.T., Ng, A., Kranich, J., Sierro, F., Yu, D., Schilter, H.C., Rolph, M.S., Mackay, F., Artis, D., Xavier, R.J., Teixeira, M.M., and Mackay, C.R. (2009). Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature* 461**,** 1282-1286.
341. Maxam, A.M., and Gilbert, W. (1977). A new method for sequencing DNA. *Proc Natl Acad Sci U S A* 74**,** 560-564.
342. Mcgeachy, M.J., Chen, Y., Tato, C.M., Laurence, A., Joyce-Shaikh, B., Blumenschein, W.M., Mcclanahan, T.K., O'shea, J.J., and Cua, D.J. (2009). The interleukin 23 receptor is essential for the terminal differentiation of interleukin 17-producing effector T helper cells in vivo. *Nat Immunol* 10**,** 314-324.
343. Mcgonagle, D., Gibbon, W., and Emery, P. (1998). Classification of inflammatory arthritis by enthesitis. *Lancet* 352**,** 1137-1140.
344. Mcintosh, K., Reed, D.E., Schneider, T., Dang, F., Keshteli, A.H., De Palma, G., Madsen, K., Bercik, P., and Vanner, S. (2017). FODMAPs alter symptoms and the metabolome of patients with IBS: a randomised controlled trial. *Gut* 66**,** 1241-1251.
345. Mcknight, D.T., Huerlimann, R., Bower, D.S., Schwarzkopf, L., Alford, R.A., and Zenger, K.R. (2019). Methods for normalizing microbiome data: An ecological perspective. *Methods in Ecology and Evolution* 10**,** 389-400.

346. Mcmurdie, P.J., and Holmes, S. (2014). Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput Biol* 10**,** e1003531.
347. Mease, P., and Goffe, B.S. (2005). Diagnosis and treatment of psoriatic arthritis. *J Am Acad Dermatol* 52**,** 1-19.
348. Meinicke, P. (2015). UProC: tools for ultra-fast protein domain classification. *Bioinformatics* 31**,** 1382-1388.
349. Meng, W., Zhu, Z., Jiang, X., Too, C.L., Uebe, S., Jagodic, M., Kockum, I., Murad, S., Ferrucci, L., Alfredsson, L., Zou, H., Klareskog, L., Feinberg, A.P., Ekstrom, T.J., Padyukov, L., and Liu, Y. (2017). DNA methylation mediates genotype and smoking interaction in the development of anti-citrullinated peptide antibody-positive rheumatoid arthritis. *Arthritis Res Ther* 19**,** 71.
350. Menni, C., Jackson, M.A., Pallister, T., Steves, C.J., Spector, T.D., and Valdes, A.M. (2017). Gut microbiome diversity and high-fibre intake are related to lower long-term weight gain. *Int J Obes (Lond)* 41**,** 1099-1105.
351. Menzel, P., Ng, K.L., and Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun* 7**,** 11257.
352. Metges, C.C. (2000). Contribution of microbial amino acids to amino acid homeostasis of the host. *J Nutr* 130**,** 1857S-1864S.
353. Metzker, M.L. (2010). Sequencing technologies - the next generation. *Nat Rev Genet* 11**,** 31-46.
354. Meyer, F., Paarmann, D., D'souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., and Edwards, R.A. (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9**,** 386.
355. Mijiyawa, M., Oniankitan, O., and Khan, M.A. (2000). Spondyloarthropathies in sub-Saharan Africa. *Curr Opin Rheumatol* 12**,** 281-286.
356. Miller, D.H., Chard, D.T., and Ciccarelli, O. (2012). Clinically isolated syndromes. *Lancet Neurol* 11**,** 157-169.
357. Miller, P.D., Bolognese, M.A., Lewiecki, E.M., Mcclung, M.R., Ding, B., Austin, M., Liu, Y., San Martin, J., and Amg Bone Loss Study, G. (2008). Effect of denosumab on bone density and turnover in postmenopausal women with low bone mass after long-term continued, discontinued, and restarting of therapy: a randomized blinded phase 2 clinical trial. *Bone* 43**,** 222-229.
358. Min Jou, W., Haegeman, G., Ysebaert, M., and Fiers, W. (1972). Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature* 237**,** 82-88.
359. Minozzi, S., Bonovas, S., Lytras, T., Pecoraro, V., Gonzalez-Lorenzo, M., Bastiampillai, A.J., Gabrielli, E.M., Lonati, A.C., Moja, L., Cinquini, M., Marino, V., Matucci, A., Milano, G.M., Tocci, G., Scarpa, R., Goletti, D., and Cantini, F. (2016). Risk of infections using anti-TNF agents in rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis: a systematic review and meta-analysis. *Expert Opin Drug Saf* 15**,** 11-34.
360. Mirza, A., and Mao-Draayer, Y. (2017). The gut microbiome and microbial translocation in multiple sclerosis. *Clin Immunol* 183**,** 213-224.
361. Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.Y., El-Gebali, S., Fraser, M.I., Gough, J., Haft, D.R., Huang, H., Letunic, I., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H., Natale, D.A., Necci, M., Nuka, G., Orengo, C., Pandurangan, A.P.,

Paysan-Lafosse, T., Pesseat, S., Potter, S.C., Qureshi, M.A., Rawlings, N.D., Redaschi, N., Richardson, L.J., Rivoire, C., Salazar, G.A., Sangrador-Vegas, A., Sigrist, C.J.A., Sillitoe, I., Sutton, G.G., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Yong, S.Y., and Finn, R.D. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res* 47**,** D351-D360.

362. Mitra, A., Raychaudhuri, S.K., and Raychaudhuri, S.P. (2012). Functional role of IL-22 in psoriatic arthritis. *Arthritis Res Ther* 14**,** R65.
363. Mizrahi-Man, O., Davenport, E.R., and Gilad, Y. (2013). Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: evaluation of effective study designs. *PLoS One* 8**,** e53608.
364. Mizuno, M., Noto, D., Kaga, N., Chiba, A., and Miyake, S. (2017). The dual role of short fatty acid chains in the pathogenesis of autoimmune disease models. *PLoS One* 12**,** e0173032.
365. Moran, N.A., and Sloan, D.B. (2015). The Hologenome Concept: Helpful or Hollow? *PLoS Biol* 13**,** e1002311.
366. Morgan, X.C., Tickle, T.L., Sokol, H., Gevers, D., Devaney, K.L., Ward, D.V., Reyes, J.A., Shah, S.A., Leleiko, N., Snapper, S.B., Bousvaros, A., Korzenik, J., Sands, B.E., Xavier, R.J., and Huttenhower, C. (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* 13**,** R79.
367. Moutsianas, L., Jostins, L., Beecham, A.H., Dilthey, A.T., Xifara, D.K., Ban, M., Shah, T.S., Patsopoulos, N.A., Alfredsson, L., Anderson, C.A., Attfield, K.E., Baranzini, S.E., Barrett, J., Binder, T.M.C., Booth, D., Buck, D., Celius, E.G., Cotsapas, C., D'alfonso, S., Dendrou, C.A., Donnelly, P., Dubois, B., Fontaine, B., Fugger, L., Goris, A., Gourraud, P.A., Graetz, C., Hemmer, B., Hillert, J., International, I.B.D.G.C., Kockum, I., Leslie, S., Lill, C.M., Martinelli-Boneschi, F., Oksenberg, J.R., Olsson, T., Oturai, A., Saarela, J., Sondergaard, H.B., Spurkland, A., Taylor, B., Winkelmann, J., Zipp, F., Haines, J.L., Pericak-Vance, M.A., Spencer, C.C.A., Stewart, G., Hafler, D.A., Ivinson, A.J., Harbo, H.F., Hauser, S.L., De Jager, P.L., Compston, A., Mccauley, J.L., Sawcer, S., and Mcvean, G. (2015). Class II HLA interactions modulate genetic risk for multiple sclerosis. *Nat Genet* 47**,** 1107-1113.
368. Mudd, A.T., Berding, K., Wang, M., Donovan, S.M., and Dilger, R.N. (2017). Serum cortisol mediates the relationship between fecal Ruminococcus and brain N-acetylaspartate in the young pig. *Gut Microbes* 8**,** 589-600.
369. Mullish, B.H., Quraishi, M.N., Segal, J.P., Mccune, V.L., Baxter, M., Marsden, G.L., Moore, D.J., Colville, A., Bhala, N., Iqbal, T.H., Settle, C., Kontkowski, G., Hart, A.L., Hawkey, P.M., Goldenberg, S.D., and Williams, H.R.T. (2018). The use of faecal microbiota transplant as treatment for recurrent or refractory Clostridium difficile infection and other potential indications: joint British Society of Gastroenterology (BSG) and Healthcare Infection Society (HIS) guidelines. *Gut* 67**,** 1920-1941.
370. Murakami, S., Kurihara, S., Koikawa, N., Nakamura, A., Aoki, K., Yosigi, H., Sawaki, K., and Ohtani, M. (2009). Effects of oral supplementation with cystine and theanine on the immune function of athletes in endurance exercise: randomized, double-blind, placebo-controlled trial. *Biosci Biotechnol Biochem* 73**,** 817-821.

371. Murali, A., Bhargava, A., and Wright, E.S. (2018). IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. *Microbiome* 6**,** 140.
372. Muraro, P.A., Pasquini, M., Atkins, H.L., Bowen, J.D., Farge, D., Fassas, A., Freedman, M.S., Georges, G.E., Gualandi, F., Hamerschlak, N., Havrdova, E., Kimiskidis, V.K., Kozak, T., Mancardi, G.L., Massacesi, L., Moraes, D.A., Nash, R.A., Pavletic, S., Ouyang, J., Rovira, M., Saiz, A., Simoes, B., Trneny, M., Zhu, L., Badoglio, M., Zhong, X., Sormani, M.P., Saccardi, R., and Multiple Sclerosis-Autologous Hematopoietic Stem Cell Transplantation Long-Term Outcomes Study, G. (2017). Long-term Outcomes After Autologous Hematopoietic Stem Cell Transplantation for Multiple Sclerosis. *JAMA Neurol* 74**,** 459-469.
373. Muto, M., Date, Y., Ichimiya, M., Moriwaki, Y., Mori, K., Kamikawaji, N., Kimura, A., Sasazuki, T., and Asagami, C. (1996). Significance of antibodies to streptococcal M protein in psoriatic arthritis and their association with HLA-A*0207. *Tissue Antigens* 48**,** 645-650.
374. Myers, A., Kay, L.J., Lynch, S.A., and Walker, D.J. (2005). Recurrence risk for psoriasis and psoriatic arthritis within sibships. *Rheumatology (Oxford)* 44**,** 773-776.
375. Nagengast, F.M., Grubben, M.J., and Van Munster, I.P. (1995). Role of bile acids in colorectal carcinogenesis. *Eur J Cancer* 31A**,** 1067-1070.
376. Nagpal, R., Mainali, R., Ahmadi, S., Wang, S., Singh, R., Kavanagh, K., Kitzman, D.W., Kushugulova, A., Marotta, F., and Yadav, H. (2018). Gut microbiome and aging: Physiological and mechanistic insights. *Nutr Healthy Aging* 4**,** 267-285.
377. Nam, J.L., Takase-Minegishi, K., Ramiro, S., Chatzidionysiou, K., Smolen, J.S., Van Der Heijde, D., Bijlsma, J.W., Burmester, G.R., Dougados, M., Scholte-Voshaar, M., Van Vollenhoven, R., and Landewe, R. (2017). Efficacy of biological disease-modifying antirheumatic drugs: a systematic literature review informing the 2016 update of the EULAR recommendations for the management of rheumatoid arthritis. *Ann Rheum Dis* 76**,** 1113-1136.
378. Nash, P., and Clegg, D.O. (2005). Psoriatic arthritis therapy: NSAIDs and traditional DMARDs. *Ann Rheum Dis* 64 Suppl 2**,** ii74-77.
379. Nastasi, C., Candela, M., Bonefeld, C.M., Geisler, C., Hansen, M., Krejsgaard, T., Biagi, E., Andersen, M.H., Brigidi, P., Odum, N., Litman, T., and Woetmann, A. (2015). The effect of short-chain fatty acids on human monocyte-derived dendritic cells. *Sci Rep* 5**,** 16148.
380. Nemeth, Z.H., Bogdanovski, D.A., Barratt-Stopper, P., Paglinco, S.R., Antonioli, L., and Rolandelli, R.H. (2017). Crohn's Disease and Ulcerative Colitis Show Unique Cytokine Profiles. *Cureus* 9**,** e1177.
381. Nettleton, J.E., Reimer, R.A., and Shearer, J. (2016). Reshaping the gut microbiota: Impact of low calorie sweeteners and the link to insulin resistance? *Physiol Behav* 164**,** 488-493.
382. Ng, S.C., Shi, H.Y., Hamidi, N., Underwood, F.E., Tang, W., Benchimol, E.I., Panaccione, R., Ghosh, S., Wu, J.C.Y., Chan, F.K.L., Sung, J.J.Y., and Kaplan, G.G. (2018). Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet* 390**,** 2769-2778.

383. Nguyen, T.T., Ung, T.T., Kim, N.H., and Jung, Y.D. (2018). Role of bile acids in colon carcinogenesis. *World J Clin Cases* 6**,** 577-588.
384. Nobel, Y.R., Cox, L.M., Kirigin, F.F., Bokulich, N.A., Yamanishi, S., Teitler, I., Chung, J., Sohn, J., Barber, C.M., Goldfarb, D.S., Raju, K., Abubucker, S., Zhou, Y., Ruiz, V.E., Li, H., Mitreva, M., Alekseyenko, A.V., Weinstock, G.M., Sodergren, E., and Blaser, M.J. (2015). Metabolic and metagenomic outcomes from early-life pulsed antibiotic treatment. *Nat Commun* 6**,** 7486.
385. Noguchi, E., Homma, Y., Kang, X., Netea, M.G., and Ma, X. (2009). A Crohn's disease-associated NOD2 mutation suppresses transcription of human IL10 by inhibiting activity of the nuclear ribonucleoprotein hnRNP-A1. *Nat Immunol* 10**,** 471-479.
386. Noordenbos, T., Yeremenko, N., Gofita, I., Van De Sande, M., Tak, P.P., Canete, J.D., and Baeten, D. (2012). Interleukin-17-positive mast cells contribute to synovial inflammation in spondylarthritis. *Arthritis Rheum* 64**,** 99-109.
387. O'keefe, S.J., Li, J.V., Lahti, L., Ou, J., Carbonero, F., Mohammed, K., Posma, J.M., Kinross, J., Wahl, E., Ruder, E., Vipperla, K., Naidoo, V., Mtshali, L., Tims, S., Puylaert, P.G., Delany, J., Krasinskas, A., Benefiel, A.C., Kaseb, H.O., Newton, K., Nicholson, J.K., De Vos, W.M., Gaskins, H.R., and Zoetendal, E.G. (2015). Fat, fibre and cancer risk in African Americans and rural Africans. *Nat Commun* 6**,** 6342.
388. O'mahony, C., Scully, P., O'mahony, D., Murphy, S., O'brien, F., Lyons, A., Sherlock, G., Macsharry, J., Kiely, B., Shanahan, F., and O'mahony, L. (2008). Commensal-induced regulatory T cells mediate protection against pathogen-stimulated NF-kappaB activation. *PLoS Pathog* 4**,** e1000112.
389. O'rielly, D.D., and Rahman, P. (2014). Genetics of psoriatic arthritis. *Best Pract Res Clin Rheumatol* 28**,** 673-685.
390. Obregon-Tito, A.J., Tito, R.Y., Metcalf, J., Sankaranarayanan, K., Clemente, J.C., Ursell, L.K., Zech Xu, Z., Van Treuren, W., Knight, R., Gaffney, P.M., Spicer, P., Lawson, P., Marin-Reyes, L., Trujillo-Villarroel, O., Foster, M., Guija-Poma, E., Troncoso-Corzo, L., Warinner, C., Ozga, A.T., and Lewis, C.M. (2015). Subsistence strategies in traditional societies distinguish gut microbiomes. *Nat Commun* 6**,** 6505.
391. Ochoa-Reparaz, J., Mielcarz, D.W., Ditrio, L.E., Burroughs, A.R., Begum-Haque, S., Dasgupta, S., Kasper, D.L., and Kasper, L.H. (2010). Central nervous system demyelinating disease protection by the human commensal Bacteroides fragilis depends on polysaccharide A expression. *J Immunol* 185**,** 4101-4108.
392. Ogura, Y., Bonen, D.K., Inohara, N., Nicolae, D.L., Chen, F.F., Ramos, R., Britton, H., Moran, T., Karaliuskas, R., Duerr, R.H., Achkar, J.P., Brant, S.R., Bayless, T.M., Kirschner, B.S., Hanauer, S.B., Nunez, G., and Cho, J.H. (2001). A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 411**,** 603-606.
393. Olsson, T., Barcellos, L.F., and Alfredsson, L. (2017). Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. *Nat Rev Neurol* 13**,** 25-36.
394. Orchard, T.R., Thiyagaraja, S., Welsh, K.I., Wordsworth, B.P., Hill Gaston, J.S., and Jewell, D.P. (2000). Clinical phenotype is related to HLA genotype in the peripheral arthropathies of inflammatory bowel disease. *Gastroenterology* 118**,** 274-278.

395. Ordas, I., Eckmann, L., Talamini, M., Baumgart, D.C., and Sandborn, W.J. (2012). Ulcerative colitis. *Lancet* 380**,** 1606-1619.
396. Ortiz, G.G., Pacheco-Moises, F.P., Macias-Islas, M.A., Flores-Alvarado, L.J., Mireles-Ramirez, M.A., Gonzalez-Renovato, E.D., Hernandez-Navarro, V.E., Sanchez-Lopez, A.L., and Alatorre-Jimenez, M.A. (2014). Role of the blood-brain barrier in multiple sclerosis. *Arch Med Res* 45**,** 687-697.
397. Otani, K., Tanigawa, T., Watanabe, T., Shimada, S., Nadatani, Y., Nagami, Y., Tanaka, F., Kamata, N., Yamagami, H., Shiba, M., Tominaga, K., Fujiwara, Y., and Arakawa, T. (2017). Microbiota Plays a Key Role in Non-Steroidal Anti-Inflammatory Drug-Induced Small Intestinal Damage. *Digestion* 95**,** 22-28.
398. Ott, S.J., Musfeldt, M., Wenderoth, D.F., Hampe, J., Brant, O., Folsch, U.R., Timmis, K.N., and Schreiber, S. (2004). Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. *Gut* 53**,** 685-693.
399. Padyukov, L., Seielstad, M., Ong, R.T., Ding, B., Ronnelid, J., Seddighzadeh, M., Alfredsson, L., Klareskog, L., and Epidemiological Investigation of Rheumatoid Arthritis Study, G. (2011). A genome-wide association study suggests contrasting associations in ACPA-positive versus ACPA-negative rheumatoid arthritis. *Ann Rheum Dis* 70**,** 259-265.
400. Palfreeman, A.C., Mcnamee, K.E., and Mccann, F.E. (2013). New developments in the management of psoriasis and psoriatic arthritis: a focus on apremilast. *Drug Des Devel Ther* 7**,** 201-210.
401. Palle, P., Monaghan, K.L., Milne, S.M., and Wan, E.C.K. (2017). Cytokine Signaling in Multiple Sclerosis and Its Therapeutic Applications. *Med Sci (Basel)* 5.
402. Palleja, A., Mikkelsen, K.H., Forslund, S.K., Kashani, A., Allin, K.H., Nielsen, T., Hansen, T.H., Liang, S., Feng, Q., Zhang, C., Pyl, P.T., Coelho, L.P., Yang, H., Wang, J., Typas, A., Nielsen, M.F., Nielsen, H.B., Bork, P., Wang, J., Vilsboll, T., Hansen, T., Knop, F.K., Arumugam, M., and Pedersen, O. (2018). Recovery of gut microbiota of healthy adults following antibiotic exposure. *Nat Microbiol* 3**,** 1255-1265.
403. Palm, N.W., De Zoete, M.R., Cullen, T.W., Barry, N.A., Stefanowski, J., Hao, L., Degnan, P.H., Hu, J., Peter, I., Zhang, W., Ruggiero, E., Cho, J.H., Goodman, A.L., and Flavell, R.A. (2014). Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell* 158**,** 1000-1010.
404. Palmer, C., Bik, E.M., Digiulio, D.B., Relman, D.A., and Brown, P.O. (2007). Development of the human infant intestinal microbiota. *PLoS Biol* 5**,** e177.
405. Pan, H., Guo, R., Ju, Y., Wang, Q., Zhu, J., Xie, Y., Zheng, Y., Li, T., Liu, Z., Lu, L., Li, F., Tong, B., Xiao, L., Xu, X., Leung, E.L., Li, R., Yang, H., Wang, J., Zhou, H., Jia, H., and Liu, L. (2019). A single bacterium restores the microbiome dysbiosis to protect bones from destruction in a rat model of rheumatoid arthritis. *Microbiome* 7**,** 107.
406. Park, J., Wang, Q., Wu, Q., Mao-Draayer, Y., and Kim, C.H. (2019). Bidirectional regulatory potentials of short-chain fatty acids and their G-protein-coupled receptors in autoimmune neuroinflammation. *Sci Rep* 9**,** 8837.
407. Parte, A.C. (2018). LPSN – List of Prokaryotic names with Standing in Nomenclature (bacterio.net), 20 years on. *International Journal of Systematic and Evolutionary Microbiology* 68**,** 1825-1829.

408. Pascal, M., Perez-Gordo, M., Caballero, T., Escribese, M.M., Lopez Longo, M.N., Luengo, O., Manso, L., Matheu, V., Seoane, E., Zamorano, M., Labrador, M., and Mayorga, C. (2018). Microbiome and Allergic Diseases. *Front Immunol* 9**,** 1584.
409. Pascale, A., Marchesi, N., Marelli, C., Coppola, A., Luzi, L., Govoni, S., Giustina, A., and Gazzaruso, C. (2018). Microbiota and metabolic diseases. *Endocrine* 61**,** 357-371.
410. Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P., Collado, M.C., Rice, B.L., Dulong, C., Morgan, X.C., Golden, C.D., Quince, C., Huttenhower, C., and Segata, N. (2019). Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* 176**,** 649-662 e620.
411. Paulson, J.N., Stine, O.C., Bravo, H.C., and Pop, M. (2013). Differential abundance analysis for microbial marker-gene surveys. *Nat Methods* 10**,** 1200-1202.
412. Peluso, I., Pallone, F., and Monteleone, G. (2006). Interleukin-12 and Th1 immune response in Crohn's disease: pathogenetic relevance and therapeutic implication. *World J Gastroenterol* 12**,** 5606-5610.
413. Penders, J., Thijs, C., Vink, C., Stelma, F.F., Snijders, B., Kummeling, I., Van Den Brandt, P.A., and Stobberingh, E.E. (2006). Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics* 118**,** 511-521.
414. Peng, Y., Leung, H.C., Yiu, S.M., and Chin, F.Y. (2011). Meta-IDBA: a de Novo assembler for metagenomic data. *Bioinformatics* 27**,** i94-101.
415. Peng, Y., Leung, H.C., Yiu, S.M., and Chin, F.Y. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28**,** 1420-1428.
416. Peng, Y., Leung, H.C.M., Yiu, S.M., and Chin, F.Y.L. (Year). "IDBA – A Practical Iterative de Bruijn Graph De Novo Assembler": Springer Berlin Heidelberg), 426-440.
417. Peschken, C.A., and Esdaile, J.M. (1999). Rheumatic diseases in North America's indigenous peoples. *Semin Arthritis Rheum* 28**,** 368-391.
418. Petersson, J., Schreiber, O., Hansson, G.C., Gendler, S.J., Velcich, A., Lundberg, J.O., Roos, S., Holm, L., and Phillipson, M. (2011). Importance and regulation of the colonic mucus barrier in a mouse model of colitis. *Am J Physiol Gastrointest Liver Physiol* 300**,** G327-333.
419. Poddubnyy, D., Hermann, K.G., Callhoff, J., Listing, J., and Sieper, J. (2014). Ustekinumab for the treatment of patients with active ankylosing spondylitis: results of a 28-week, prospective, open-label, proof-of-concept study (TOPAS). *Ann Rheum Dis* 73**,** 817-823.
420. Popkin, B.M. (1999). Urbanization, Lifestyle Changes and the Nutrition Transition. *World Development* 27**,** 1905-1916.
421. Powrie, F., Leach, M.W., Mauze, S., Caddle, L.B., and Coffman, R.L. (1993). Phenotypically distinct subsets of CD4+ T cells induce or protect from chronic intestinal inflammation in C. B-17 scid mice. *Int Immunol* 5**,** 1461-1471.
422. Prantera, C., and Scribano, M.L. (2009). Antibiotics and probiotics in inflammatory bowel disease: why, when, and how. *Curr Opin Gastroenterol* 25**,** 329-333.

423. Pytlik, M., Folwarczna, J., and Janiec, W. (2004). Effects of doxycycline on mechanical properties of bones in rats with ovariectomy-induced osteopenia. *Calcif Tissue Int* 75**,** 225-230.
424. Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D.R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H.B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Dore, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Meta, H.I.T.C., Bork, P., Ehrlich, S.D., and Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464**,** 59-65.
425. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glockner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41**,** D590-596.
426. Radner, H., Lesperance, T., Accortt, N.A., and Solomon, D.H. (2017). Incidence and Prevalence of Cardiovascular Risk Factors Among Patients With Rheumatoid Arthritis, Psoriasis, or Psoriatic Arthritis. *Arthritis Care Res (Hoboken)* 69**,** 1510-1518.
427. Rajendran, N., and Kumar, D. (2010). Role of diet in the management of inflammatory bowel disease. *World J Gastroenterol* 16**,** 1442-1448.
428. Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S., and Medzhitov, R. (2004). Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell* 118**,** 229-241.
429. Ralston, S.H., Uitterlinden, A.G., Brandi, M.L., Balcells, S., Langdahl, B.L., Lips, P., Lorenc, R., Obermayer-Pietsch, B., Scollen, S., Bustamante, M., Husted, L.B., Carey, A.H., Diez-Perez, A., Dunning, A.M., Falchetti, A., Karczmarewicz, E., Kruk, M., Van Leeuwen, J.P., Van Meurs, J.B., Mangion, J., Mcguigan, F.E., Mellibovsky, L., Del Monte, F., Pols, H.A., Reeve, J., Reid, D.M., Renner, W., Rivadeneira, F., Van Schoor, N.M., Sherlock, R.E., Ioannidis, J.P., and Investigators, G. (2006). Large-scale evidence for the effect of the COLIA1 Sp1 polymorphism on osteoporosis outcomes: the GENOMOS study. *PLoS Med* 3**,** e90.
430. Ramos, M.I., Teunissen, M.B., Helder, B., Aarrass, S., De Hair, M.J., Van Kuijk, A.W., Gerlag, D.M., Tak, P.P., and Lebre, M.C. (2016). Reduced CLEC9A expression in synovial tissue of psoriatic arthritis patients after adalimumab therapy. *Rheumatology (Oxford)* 55**,** 1575-1584.
431. Rath, H.C., Herfarth, H.H., Ikeda, J.S., Grenther, W.B., Hamm, T.E., Jr., Balish, E., Taurog, J.D., Hammer, R.E., Wilson, K.H., and Sartor, R.B. (1996). Normal luminal bacteria, especially Bacteroides species, mediate chronic colitis, gastritis, and arthritis in HLA-B27/human beta2 microglobulin transgenic rats. *J Clin Invest* 98**,** 945-953.
432. Rawla, P., Sunkara, T., and Raj, J.P. (2018). Role of biologics and biosimilars in inflammatory bowel disease: current trends and future perspectives. *J Inflamm Res* 11**,** 215-226.
433. Redlich, K., Hayer, S., Ricci, R., David, J.P., Tohidast-Akrad, M., Kollias, G., Steiner, G., Smolen, J.S., Wagner, E.F., and Schett, G. (2002). Osteoclasts are

essential for TNF-alpha-mediated joint destruction. *J Clin Invest* 110**,** 1419-1427.
434. Rehakova, Z., Capkova, J., Stepankova, R., Sinkora, J., Louzecka, A., Ivanyi, P., and Weinreich, S. (2000). Germ-free mice do not develop ankylosing enthesopathy, a spontaneous joint disease. *Hum Immunol* 61**,** 555-558.
435. Rideout, J.R., He, Y., Navas-Molina, J.A., Walters, W.A., Ursell, L.K., Gibbons, S.M., Chase, J., Mcdonald, D., Gonzalez, A., Robbins-Pianka, A., Clemente, J.C., Gilbert, J.A., Huse, S.M., Zhou, H.W., Knight, R., and Caporaso, J.G. (2014). Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. *PeerJ* 2**,** e545.
436. Ridlon, J.M., Kang, D.J., Hylemon, P.B., and Bajaj, J.S. (2014). Bile acids and the gut microbiome. *Curr Opin Gastroenterol* 30**,** 332-338.
437. Riesenfeld, C.S., Schloss, P.D., and Handelsman, J. (2004). Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet* 38**,** 525-552.
438. Rioux, J.D., Xavier, R.J., Taylor, K.D., Silverberg, M.S., Goyette, P., Huett, A., Green, T., Kuballa, P., Barmada, M.M., Datta, L.W., Shugart, Y.Y., Griffiths, A.M., Targan, S.R., Ippoliti, A.F., Bernard, E.J., Mei, L., Nicolae, D.L., Regueiro, M., Schumm, L.P., Steinhart, A.H., Rotter, J.I., Duerr, R.H., Cho, J.H., Daly, M.J., and Brant, S.R. (2007). Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet* 39**,** 596-604.
439. Ritchlin, C.T., Colbert, R.A., and Gladman, D.D. (2017). Psoriatic Arthritis. *N Engl J Med* 376**,** 957-970.
440. Robinson, M.D., Mccarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26**,** 139-140.
441. Rodriguez, J.M., Murphy, K., Stanton, C., Ross, R.P., Kober, O.I., Juge, N., Avershina, E., Rudi, K., Narbad, A., Jenmalm, M.C., Marchesi, J.R., and Collado, M.C. (2015). The composition of the gut microbiota throughout life, with an emphasis on early life. *Microb Ecol Health Dis* 26**,** 26050.
442. Rogier, R., Ederveen, T.H.A., Boekhorst, J., Wopereis, H., Scher, J.U., Manasson, J., Frambach, S., Knol, J., Garssen, J., Van Der Kraan, P.M., Koenders, M.I., Van Den Berg, W.B., Van Hijum, S., and Abdollahi-Roodsaz, S. (2017). Aberrant intestinal microbiota due to IL-1 receptor antagonist deficiency promotes IL-17- and TLR4-dependent arthritis. *Microbiome* 5**,** 63.
443. Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M., and Nyren, P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem* 242**,** 84-89.
444. Rook, G.A., Lowry, C.A., and Raison, C.L. (2013). Microbial 'Old Friends', immunoregulation and stress resilience. *Evol Med Public Health* 2013**,** 46-64.
445. Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P.I., Godneva, A., Kalka, I.N., Bar, N., Shilo, S., Lador, D., Vila, A.V., Zmora, N., Pevsner-Fischer, M., Israeli, D., Kosower, N., Malka, G., Wolf, B.C., Avnit-Sagi, T., Lotan-Pompan, M., Weinberger, A., Halpern, Z., Carmi, S., Fu, J., Wijmenga, C., Zhernakova, A., Elinav, E., and Segal, E. (2018). Environment dominates over host genetics in shaping human gut microbiota. *Nature* 555**,** 210-215.

446. Rovira, A., Wattjes, M.P., Tintore, M., Tur, C., Yousry, T.A., Sormani, M.P., De Stefano, N., Filippi, M., Auger, C., Rocca, M.A., Barkhof, F., Fazekas, F., Kappos, L., Polman, C., Miller, D., Montalban, X., and Group, M.S. (2015). Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis-clinical implementation in the diagnostic process. *Nat Rev Neurol* 11**,** 471-482.
447. Rowland, I., Gibson, G., Heinken, A., Scott, K., Swann, J., Thiele, I., and Tuohy, K. (2018). Gut microbiota functions: metabolism of nutrients and other food components. *Eur J Nutr* 57**,** 1-24.
448. Rudwaleit, M., Haibel, H., Baraliakos, X., Listing, J., Marker-Hermann, E., Zeidler, H., Braun, J., and Sieper, J. (2009a). The early disease stage in axial spondylarthritis: results from the German Spondyloarthritis Inception Cohort. *Arthritis Rheum* 60**,** 717-727.
449. Rudwaleit, M., Khan, M.A., and Sieper, J. (2005). The challenge of diagnosis and classification in early ankylosing spondylitis: do we need new criteria? *Arthritis Rheum* 52**,** 1000-1008.
450. Rudwaleit, M., Van Der Heijde, D., Landewe, R., Akkoc, N., Brandt, J., Chou, C.T., Dougados, M., Huang, F., Gu, J., Kirazli, Y., Van Den Bosch, F., Olivieri, I., Roussou, E., Scarpato, S., Sorensen, I.J., Valle-Onate, R., Weber, U., Wei, J., and Sieper, J. (2011). The Assessment of SpondyloArthritis International Society classification criteria for peripheral spondyloarthritis and for spondyloarthritis in general. *Ann Rheum Dis* 70**,** 25-31.
451. Rudwaleit, M., Van Der Heijde, D., Landewe, R., Listing, J., Akkoc, N., Brandt, J., Braun, J., Chou, C.T., Collantes-Estevez, E., Dougados, M., Huang, F., Gu, J., Khan, M.A., Kirazli, Y., Maksymowych, W.P., Mielants, H., Sorensen, I.J., Ozgocmen, S., Roussou, E., Valle-Onate, R., Weber, U., Wei, J., and Sieper, J. (2009b). The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. *Ann Rheum Dis* 68**,** 777-783.
452. Rusk, N. (2011). Torrents of sequence. *Nature Methods* 8**,** 44-44.
453. Russmueller, G., Seemann, R., Weiss, K., Stadler, V., Speiss, M., Perisanidis, C., Fuereder, T., Willinger, B., Sulzbacher, I., and Steininger, C. (2016). The association of medication-related osteonecrosis of the jaw with Actinomyces spp. infection. *Sci Rep* 6**,** 31604.
454. Ryle, A.P., Sanger, F., Smith, L.F., and Kitai, R. (1955). The disulphide bonds of insulin. *Biochem J* 60**,** 541-556.
455. Salim, S.Y., and Soderholm, J.D. (2011). Importance of disrupted intestinal barrier in inflammatory bowel diseases. *Inflamm Bowel Dis* 17**,** 362-381.
456. Salyers, A.A., Palmer, J.K., and Wilkins, T.D. (1978). Degradation of polysaccharides by intestinal bacterial enzymes. *Am J Clin Nutr* 31**,** S128-S130.
457. Salzberg, S.L., Delcher, A.L., Kasif, S., and White, O. (1998). Microbial gene identification using interpolated Markov models. *Nucleic Acids Res* 26**,** 544-548.
458. Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F., and Petersen, G.B. (1982). Nucleotide sequence of bacteriophage lambda DNA. *J Mol Biol* 162**,** 729-773.
459. Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74**,** 5463-5467.
460. Sankowski, A.J., Lebkowska, U.M., Cwikla, J., Walecka, I., and Walecki, J. (2013). The comparison of efficacy of different imaging techniques

(conventional radiography, ultrasonography, magnetic resonance) in assessment of wrist joints and metacarpophalangeal joints in patients with psoriatic arthritis. *Pol J Radiol* 78**,** 18-29.
461. Satsangi, J., Silverberg, M.S., Vermeire, S., and Colombel, J.F. (2006). The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut* 55**,** 749-753.
462. Savage, D.C. (1977). Microbial ecology of the gastrointestinal tract. *Annu Rev Microbiol* 31**,** 107-133.
463. Schadt, E.E., Turner, S., and Kasarskis, A. (2010). A window into third-generation sequencing. *Hum Mol Genet* 19**,** R227-240.
464. Schaubeck, M., Clavel, T., Calasan, J., Lagkouvardos, I., Haange, S.B., Jehmlich, N., Basic, M., Dupont, A., Hornef, M., Von Bergen, M., Bleich, A., and Haller, D. (2016). Dysbiotic gut microbiota causes transmissible Crohn's disease-like ileitis independent of failure in antimicrobial defence. *Gut* 65**,** 225-237.
465. Schepers, E., Meert, N., Glorieux, G., Goeman, J., Van Der Eycken, J., and Vanholder, R. (2007). P-cresylsulphate, the main in vivo metabolite of p-cresol, activates leucocyte free radical production. *Nephrol Dial Transplant* 22**,** 592-596.
466. Scher, J.U., Sczesnak, A., Longman, R.S., Segata, N., Ubeda, C., Bielski, C., Rostron, T., Cerundolo, V., Pamer, E.G., Abramson, S.B., Huttenhower, C., and Littman, D.R. (2013). Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. *Elife* 2**,** e01202.
467. Scher, J.U., Ubeda, C., Artacho, A., Attur, M., Isaac, S., Reddy, S.M., Marmon, S., Neimann, A., Brusca, S., Patel, T., Manasson, J., Pamer, E.G., Littman, D.R., and Abramson, S.B. (2015). Decreased bacterial diversity characterizes the altered gut microbiota in patients with psoriatic arthritis, resembling dysbiosis in inflammatory bowel disease. *Arthritis Rheumatol* 67**,** 128-139.
468. Schett, G., and Gravallese, E. (2012). Bone erosion in rheumatoid arthritis: mechanisms, diagnosis and treatment. *Nat Rev Rheumatol* 8**,** 656-664.
469. Schirmer, M., Franzosa, E.A., Lloyd-Price, J., Mciver, L.J., Schwager, R., Poon, T.W., Ananthakrishnan, A.N., Andrews, E., Barron, G., Lake, K., Prasad, M., Sauk, J., Stevens, B., Wilson, R.G., Braun, J., Denson, L.A., Kugathasan, S., Mcgovern, D.P.B., Vlamakis, H., Xavier, R.J., and Huttenhower, C. (2018). Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat Microbiol* 3**,** 337-346.
470. Schirmer, M., Ijaz, U.Z., D'amore, R., Hall, N., Sloan, W.T., and Quince, C. (2015). Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res* 43**,** e37.
471. Schirmer, M., Smeekens, S.P., Vlamakis, H., Jaeger, M., Oosting, M., Franzosa, E.A., Ter Horst, R., Jansen, T., Jacobs, L., Bonder, M.J., Kurilshikov, A., Fu, J., Joosten, L.a.B., Zhernakova, A., Huttenhower, C., Wijmenga, C., Netea, M.G., and Xavier, R.J. (2016). Linking the Human Gut Microbiome to Inflammatory Cytokine Production Capacity. *Cell* 167**,** 1125-1136 e1128.
472. Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., and Weber, C.F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for

describing and comparing microbial communities. *Appl Environ Microbiol* 75**,** 7537-7541.
473. Schlosstein, L., Terasaki, P.I., Bluestone, R., and Pearson, C.M. (1973). High association of an HL-A antigen, W27, with ankylosing spondylitis. *N Engl J Med* 288**,** 704-706.
474. Schmidt, T.S., Matias Rodrigues, J.F., and Von Mering, C. (2014). Ecological consistency of SSU rRNA-based operational taxonomic units at a global scale. *PLoS Comput Biol* 10**,** e1003594.
475. Schnorr, S.L., Candela, M., Rampelli, S., Centanni, M., Consolandi, C., Basaglia, G., Turroni, S., Biagi, E., Peano, C., Severgnini, M., Fiori, J., Gotti, R., De Bellis, G., Luiselli, D., Brigidi, P., Mabulla, A., Marlowe, F., Henry, A.G., and Crittenden, A.N. (2014). Gut microbiome of the Hadza hunter-gatherers. *Nat Commun* 5**,** 3654.
476. Scholz, J.L., Diaz, A., Riley, R.L., Cancro, M.P., and Frasca, D. (2013). A comparative review of aging and B cell function in mice and humans. *Curr Opin Immunol* 25**,** 504-510.
477. Schwan, A., Sjolin, S., Trottestam, U., and Aronsson, B. (1983). Relapsing clostridium difficile enterocolitis cured by rectal infusion of homologous faeces. *Lancet* 2**,** 845.
478. Schwenzer, N.F., Kotter, I., Henes, J.C., Schraml, C., Fritz, J., Claussen, C.D., and Horger, M. (2010). The role of dynamic contrast-enhanced MRI in the differential diagnosis of psoriatic and rheumatoid arthritis. *AJR Am J Roentgenol* 194**,** 715-720.
479. Scott, D.L. (2004). Radiological progression in established rheumatoid arthritis. *J Rheumatol Suppl* 69**,** 55-65.
480. Sefik, E., Geva-Zatorsky, N., Oh, S., Konnikova, L., Zemmour, D., Mcguire, A.M., Burzyn, D., Ortiz-Lopez, A., Lobera, M., Yang, J., Ghosh, S., Earl, A., Snapper, S.B., Jupp, R., Kasper, D., Mathis, D., and Benoist, C. (2015). MUCOSAL IMMUNOLOGY. Individual intestinal symbionts induce a distinct population of RORgamma(+) regulatory T cells. *Science* 349**,** 993-997.
481. Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* 9**,** 811-814.
482. Sender, R., Fuchs, S., and Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol* 14**,** e1002533.
483. Shah, S.C., Khalili, H., Gower-Rousseau, C., Olen, O., Benchimol, E.I., Lynge, E., Nielsen, K.R., Brassard, P., Vutcovici, M., Bitton, A., Bernstein, C.N., Leddin, D., Tamim, H., Stefansson, T., Loftus, E.V., Jr., Moum, B., Tang, W., Ng, S.C., Gearry, R., Sincic, B., Bell, S., Sands, B.E., Lakatos, P.L., Vegh, Z., Ott, C., Kaplan, G.G., Burisch, J., and Colombel, J.F. (2018). Sex-Based Differences in Incidence of Inflammatory Bowel Diseases-Pooled Analysis of Population-Based Studies From Western Countries. *Gastroenterology* 155**,** 1079-1089 e1073.
484. Shahi, S.K., Freedman, S.N., and Mangalam, A.K. (2017). Gut microbiome in multiple sclerosis: The players involved and the roles they play. *Gut Microbes* 8**,** 607-615.
485. Shapiro, J., Cohen, N.A., Shalev, V., Uzan, A., Koren, O., and Maharshak, N. (2019). Psoriatic patients have a distinct structural and functional fecal microbiota compared with controls. *J Dermatol* 46**,** 595-603.

486. Sharma, R., Schumacher, U., Ronaasen, V., and Coates, M. (1995). Rat intestinal mucosal responses to a microbial flora and different diets. *Gut* 36**,** 209-214.
487. Shea, B., Wells, G., Cranney, A., Zytaruk, N., Robinson, V., Griffith, L., Ortiz, Z., Peterson, J., Adachi, J., Tugwell, P., Guyatt, G., Osteoporosis Methodology, G., and The Osteoporosis Research Advisory, G. (2002). Meta-analyses of therapies for postmenopausal osteoporosis. VII. Meta-analysis of calcium supplementation for the prevention of postmenopausal osteoporosis. *Endocr Rev* 23**,** 552-559.
488. Sherlock, J.P., Joyce-Shaikh, B., Turner, S.P., Chao, C.C., Sathe, M., Grein, J., Gorman, D.M., Bowman, E.P., Mcclanahan, T.K., Yearley, J.H., Eberl, G., Buckley, C.D., Kastelein, R.A., Pierce, R.H., Laface, D.M., and Cua, D.J. (2012). IL-23 induces spondyloarthropathy by acting on ROR-gammat+ CD3+CD4-CD8- entheseal resident T cells. *Nat Med* 18**,** 1069-1076.
489. Sieper, J., Braun, J., Dougados, M., and Baeten, D. (2015). Axial spondyloarthritis. *Nat Rev Dis Primers* 1**,** 15013.
490. Sieper, J., Braun, J., Rudwaleit, M., Boonen, A., and Zink, A. (2002). Ankylosing spondylitis: an overview. *Ann Rheum Dis* 61 Suppl 3**,** iii8-18.
491. Sieper, J., and Van Der Heijde, D. (2013). Review: Nonradiographic axial spondyloarthritis: new definition of an old disease? *Arthritis Rheum* 65**,** 543-551.
492. Silman, A.J., Macgregor, A.J., Thomson, W., Holligan, S., Carthy, D., Farhan, A., and Ollier, W.E. (1993). Twin concordance rates for rheumatoid arthritis: results from a nationwide study. *Br J Rheumatol* 32**,** 903-907.
493. Silman, A.J., and Pearson, J.E. (2002). Epidemiology and genetics of rheumatoid arthritis. *Arthritis Res* 4 Suppl 3**,** S265-272.
494. Simpson, E.H. (1949). Measurement of Diversity. *Nature* 163**,** 688-688.
495. Simpson, J.T., Workman, R.E., Zuzarte, P.C., David, M., Dursi, L.J., and Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods* 14**,** 407-410.
496. Sinkorova, Z., Capkova, J., Niederlova, J., Stepankova, R., and Sinkora, J. (2008). Commensal intestinal bacterial strains trigger ankylosing enthesopathy of the ankle in inbred B10.BR (H-2(k)) male mice. *Hum Immunol* 69**,** 845-850.
497. Siris, E.S., Gehlbach, S., Adachi, J.D., Boonen, S., Chapurlat, R.D., Compston, J.E., Cooper, C., Delmas, P., Diez-Perez, A., Hooven, F.H., Lacroix, A.Z., Netelenbos, J.C., Pfeilschifter, J., Rossini, M., Roux, C., Saag, K.G., Sambrook, P., Silverman, S., Watts, N.B., Wyman, A., and Greenspan, S.L. (2011). Failure to perceive increased risk of fracture in women 55 years and older: the Global Longitudinal Study of Osteoporosis in Women (GLOW). *Osteoporos Int* 22**,** 27-35.
498. Sjogren, K., Engdahl, C., Henning, P., Lerner, U.H., Tremaroli, V., Lagerquist, M.K., Backhed, F., and Ohlsson, C. (2012). The gut microbiota regulates bone mass in mice. *J Bone Miner Res* 27**,** 1357-1367.
499. Sjogren, Y.M., Jenmalm, M.C., Bottcher, M.F., Bjorksten, B., and Sverremark-Ekstrom, E. (2009). Altered early infant gut microbiota in children developing allergy up to 5 years of age. *Clin Exp Allergy* 39**,** 518-526.
500. Skrypnik, K., and Suliburska, J. (2018). Association between the gut microbiota and mineral metabolism. *J Sci Food Agric* 98**,** 2449-2460.

501. Smith, P.M., Howitt, M.R., Panikov, N., Michaud, M., Gallini, C.A., Bohlooly, Y.M., Glickman, J.N., and Garrett, W.S. (2013). The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. *Science* 341**,** 569-573.
502. Smolen, J.S., Emery, P., Fleischmann, R., Van Vollenhoven, R.F., Pavelka, K., Durez, P., Guerette, B., Kupper, H., Redden, L., Arora, V., and Kavanaugh, A. (2014). Adjustment of therapy in rheumatoid arthritis on the basis of achievement of stable low disease activity with adalimumab plus methotrexate or methotrexate alone: the randomised controlled OPTIMA trial. *Lancet* 383**,** 321-332.
503. Smolen, J.S., Landewe, R., Bijlsma, J., Burmester, G., Chatzidionysiou, K., Dougados, M., Nam, J., Ramiro, S., Voshaar, M., Van Vollenhoven, R., Aletaha, D., Aringer, M., Boers, M., Buckley, C.D., Buttgereit, F., Bykerk, V., Cardiel, M., Combe, B., Cutolo, M., Van Eijk-Hustings, Y., Emery, P., Finckh, A., Gabay, C., Gomez-Reino, J., Gossec, L., Gottenberg, J.E., Hazes, J.M.W., Huizinga, T., Jani, M., Karateev, D., Kouloumas, M., Kvien, T., Li, Z., Mariette, X., Mcinnes, I., Mysler, E., Nash, P., Pavelka, K., Poor, G., Richez, C., Van Riel, P., Rubbert-Roth, A., Saag, K., Da Silva, J., Stamm, T., Takeuchi, T., Westhovens, R., De Wit, M., and Van Der Heijde, D. (2017). EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. *Ann Rheum Dis* 76**,** 960-977.
504. Sokka, T., Toloza, S., Cutolo, M., Kautiainen, H., Makinen, H., Gogus, F., Skakic, V., Badsha, H., Peets, T., Baranauskaite, A., Geher, P., Ujfalussy, I., Skopouli, F.N., Mavrommati, M., Alten, R., Pohl, C., Sibilia, J., Stancati, A., Salaffi, F., Romanowski, W., Zarowny-Wierzbinska, D., Henrohn, D., Bresnihan, B., Minnock, P., Knudsen, L.S., Jacobs, J.W., Calvo-Alen, J., Lazovskis, J., Pinheiro Gda, R., Karateev, D., Andersone, D., Rexhepi, S., Yazici, Y., Pincus, T., and Group, Q.-R. (2009). Women, men, and rheumatoid arthritis: analyses of disease activity, disease characteristics, and treatments in the QUEST-RA study. *Arthritis Res Ther* 11**,** R7.
505. Sokolove, J., Wagner, C.A., Lahey, L.J., Sayles, H., Duryee, M.J., Reimold, A.M., Kerr, G., Robinson, W.H., Cannon, G.W., Thiele, G.M., and Mikuls, T.R. (2016). Increased inflammation and disease activity among current cigarette smokers with rheumatoid arthritis: a cross-sectional analysis of US veterans. *Rheumatology (Oxford)* 55**,** 1969-1977.
506. Song, H.J., Grant, I., Rotondo, D., Mohede, I., Sattar, N., Heys, S.D., and Wahle, K.W. (2005). Effect of CLA supplementation on immune function in young healthy volunteers. *Eur J Clin Nutr* 59**,** 508-517.
507. Sood, A., Midha, V., Makharia, G.K., Ahuja, V., Singal, D., Goswami, P., and Tandon, R.K. (2009). The probiotic preparation, VSL#3 induces remission in patients with mild-to-moderately active ulcerative colitis. *Clin Gastroenterol Hepatol* 7**,** 1202-1209, 1209 e1201.
508. Spellerberg, I.F., and Fedor, P.J. (2003). A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon–Wiener' Index. *Global Ecology and Biogeography* 12**,** 177-179.
509. Srivastava, R.K., Dar, H.Y., and Mishra, P.K. (2018). Immunoporosis: Immunology of Osteoporosis-Role of T Cells. *Front Immunol* 9**,** 657.

510. Staley, C., Weingarden, A.R., Khoruts, A., and Sadowsky, M.J. (2017). Interaction of gut microbiota with bile acid metabolism and its influence on disease states. *Appl Microbiol Biotechnol* 101**,** 47-64.
511. Stark, P.L., and Lee, A. (1982). The microbial ecology of the large bowel of breast-fed and formula-fed infants during the first year of life. *J Med Microbiol* 15**,** 189-203.
512. Stebbings, S.M., Taylor, C., Tannock, G.W., Baird, M.A., and Highton, J. (2009). The immune response to autologous bacteroides in ankylosing spondylitis is characterized by reduced interleukin 10 production. *J Rheumatol* 36**,** 797-800.
513. Stecher, B., and Hardt, W.D. (2008). The role of microbiota in infectious disease. *Trends Microbiol* 16**,** 107-114.
514. Stoll, M.L., Kumar, R., Morrow, C.D., Lefkowitz, E.J., Cui, X., Genin, A., Cron, R.Q., and Elson, C.O. (2014). Altered microbiota associated with abnormal humoral immune responses to commensal organisms in enthesitis-related arthritis. *Arthritis Res Ther* 16**,** 486.
515. Strachan, D.P. (1989). Hay fever, hygiene, and household size. *BMJ* 299**,** 1259-1260.
516. Stuart, P.E., Nair, R.P., Tsoi, L.C., Tejasvi, T., Das, S., Kang, H.M., Ellinghaus, E., Chandran, V., Callis-Duffin, K., Ike, R., Li, Y., Wen, X., Enerback, C., Gudjonsson, J.E., Koks, S., Kingo, K., Esko, T., Mrowietz, U., Reis, A., Wichmann, H.E., Gieger, C., Hoffmann, P., Nothen, M.M., Winkelmann, J., Kunz, M., Moreta, E.G., Mease, P.J., Ritchlin, C.T., Bowcock, A.M., Krueger, G.G., Lim, H.W., Weidinger, S., Weichenthal, M., Voorhees, J.J., Rahman, P., Gregersen, P.K., Franke, A., Gladman, D.D., Abecasis, G.R., and Elder, J.T. (2015). Genome-wide Association Analysis of Psoriatic Arthritis and Cutaneous Psoriasis Reveals Differences in Their Genetic Architecture. *Am J Hum Genet* 97**,** 816-836.
517. Sucur, A., Jajic, Z., Artukovic, M., Matijasevic, M.I., Anic, B., Flegar, D., Markotic, A., Kelava, T., Ivcevic, S., Kovacic, N., Katavic, V., and Grcevic, D. (2017). Chemokine signals are crucial for enhanced homing and differentiation of circulating osteoclast progenitor cells. *Arthritis Res Ther* 19**,** 142.
518. Sudo, N., Sawamura, S., Tanaka, K., Aiba, Y., Kubo, C., and Koga, Y. (1997). The requirement of intestinal bacterial flora for the development of an IgE production system fully susceptible to oral tolerance induction. *J Immunol* 159**,** 1739-1745.
519. Sultan, A.A., Mallen, C., Muller, S., Hider, S., Scott, I., Helliwell, T., and Hall, L.J. (2019). Antibiotic use and the risk of rheumatoid arthritis: a population-based case-control study. *BMC Med* 17**,** 154.
520. Suzek, B.E., Wang, Y., Huang, H., Mcgarvey, P.B., Wu, C.H., and Uniprot, C. (2015). UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* 31**,** 926-932.
521. Szulc, P., Seeman, E., Duboeuf, F., Sornay-Rendu, E., and Delmas, P.D. (2006). Bone fragility: failure of periosteal apposition to compensate for increased endocortical resorption in postmenopausal women. *J Bone Miner Res* 21**,** 1856-1863.
522. Taams, L.S., Steel, K.J.A., Srenathan, U., Burns, L.A., and Kirkham, B.W. (2018). IL-17 in the immunopathogenesis of spondyloarthritis. *Nat Rev Rheumatol* 14**,** 453-466.

523. Tamburini, S., Shen, N., Wu, H.C., and Clemente, J.C. (2016). The microbiome in early life: implications for health outcomes. *Nat Med* 22**,** 713-722.
524. Tang, W.H., Wang, Z., Levison, B.S., Koeth, R.A., Britt, E.B., Fu, X., Wu, Y., and Hazen, S.L. (2013). Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N Engl J Med* 368**,** 1575-1584.
525. Tankou, S.K., Regev, K., Healy, B.C., Tjon, E., Laghi, L., Cox, L.M., Kivisakk, P., Pierre, I.V., Hrishikesh, L., Gandhi, R., Cook, S., Glanz, B., Stankiewicz, J., and Weiner, H.L. (2018). A probiotic modulates the microbiome and immunity in multiple sclerosis. *Ann Neurol* 83**,** 1147-1161.
526. Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 28**,** 33-36.
527. Taylor, W., Gladman, D., Helliwell, P., Marchesoni, A., Mease, P., Mielants, H., and Group, C.S. (2006). Classification criteria for psoriatic arthritis: development of new criteria from a large international study. *Arthritis Rheum* 54**,** 2665-2673.
528. Thompson, A.J., Banwell, B.L., Barkhof, F., Carroll, W.M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M.S., Fujihara, K., Galetta, S.L., Hartung, H.P., Kappos, L., Lublin, F.D., Marrie, R.A., Miller, A.E., Miller, D.H., Montalban, X., Mowry, E.M., Sorensen, P.S., Tintore, M., Traboulsee, A.L., Trojano, M., Uitdehaag, B.M.J., Vukusic, S., Waubant, E., Weinshenker, B.G., Reingold, S.C., and Cohen, J.A. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol* 17**,** 162-173.
529. Ticinesi, A., Milani, C., Lauretani, F., Nouvenne, A., Mancabelli, L., Lugli, G.A., Turroni, F., Duranti, S., Mangifesta, M., Viappiani, A., Ferrario, C., Maggio, M., Ventura, M., and Meschi, T. (2017). Gut microbiota composition is associated with polypharmacy in elderly hospitalized patients. *Sci Rep* 7**,** 11102.
530. Tintore, M., Vidal-Jordana, A., and Sastre-Garriga, J. (2019). Treatment of multiple sclerosis - success from bench to bedside. *Nat Rev Neurol* 15**,** 53-58.
531. Tissier, H. (1900). *Recherches sur la flore intestinale des nourrissons (état normal et pathologique).* Paris: Georges Carré et C. Naud.
532. Tobon, G.J., Youinou, P., and Saraux, A. (2010). The environment, geo-epidemiology, and autoimmune disease: Rheumatoid arthritis. *J Autoimmun* 35**,** 10-14.
533. Toes, R.E., and Van Der Woude, D. (2011). ACPA (anti-citrullinated protein antibodies) and rheumatoid arthritis. *Acta Reumatol Port* 36**,** 205-207.
534. Torre Alonso, J.C., Rodriguez Perez, A., Arribas Castrillo, J.M., Ballina Garcia, J., Riestra Noriega, J.L., and Lopez Larrea, C. (1991). Psoriatic arthritis (PA): a clinical, immunological and radiological study of 180 patients. *Br J Rheumatol* 30**,** 245-250.
535. Tremelling, M., Cummings, F., Fisher, S.A., Mansfield, J., Gwilliam, R., Keniry, A., Nimmo, E.R., Drummond, H., Onnie, C.M., Prescott, N.J., Sanderson, J., Bredin, F., Berzuini, C., Forbes, A., Lewis, C.M., Cardon, L., Deloukas, P., Jewell, D., Mathew, C.G., Parkes, M., and Satsangi, J. (2007). IL23R variation determines susceptibility but not disease phenotype in inflammatory bowel disease. *Gastroenterology* 132**,** 1657-1664.
536. Tricon, S., Burdge, G.C., Kew, S., Banerjee, T., Russell, J.J., Grimble, R.F., Williams, C.M., Calder, P.C., and Yaqoob, P. (2004). Effects of cis-9,trans-11

and trans-10,cis-12 conjugated linoleic acid on immune cell function in healthy humans. *Am J Clin Nutr* 80**,** 1626-1633.
537. Trifan, A., Stanciu, C., Girleanu, I., Stoica, O.C., Singeap, A.M., Maxim, R., Chiriac, S.A., Ciobica, A., and Boiculese, L. (2017). Proton pump inhibitors therapy and risk of Clostridium difficile infection: Systematic review and meta-analysis. *World J Gastroenterol* 23**,** 6500-6515.
538. Truong, D.T., Franzosa, E.A., Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., and Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods* 12**,** 902-903.
539. Truong, D.T., Tett, A., Pasolli, E., Huttenhower, C., and Segata, N. (2017). Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Res* 27**,** 626-638.
540. Tun, H.M., Konya, T., Takaro, T.K., Brook, J.R., Chari, R., Field, C.J., Guttman, D.S., Becker, A.B., Mandhane, P.J., Turvey, S.E., Subbarao, P., Sears, M.R., Scott, J.A., Kozyrskyj, A.L., and Investigators, C.S. (2017). Exposure to household furry pets influences the gut microbiota of infant at 3-4 months following various birth scenarios. *Microbiome* 5**,** 40.
541. Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R., and Gordon, J.I. (2007). The human microbiome project. *Nature* 449**,** 804-810.
542. Turner, J.R. (2009). Intestinal mucosal barrier function in health and disease. *Nat Rev Immunol* 9**,** 799-809.
543. Turner, W.R., Nakamura, T., and Dinetti, M. (2004). Global Urbanization and the Separation of Humans from Nature. *BioScience* 54**,** 585-590.
544. Uejima, M., Kinouchi, T., Kataoka, K., Hiraoka, I., and Ohnishi, Y. (1996). Role of intestinal bacteria in ileal ulcer formation in rats treated with a nonsteroidal antiinflammatory drug. *Microbiol Immunol* 40**,** 553-560.
545. Ueno, A., Jeffery, L., Kobayashi, T., Hibi, T., Ghosh, S., and Jijon, H. (2018). Th17 plasticity and its relevance to inflammatory bowel disease. *J Autoimmun* 87**,** 38-49.
546. Uitterlinden, A.G., Ralston, S.H., Brandi, M.L., Carey, A.H., Grinberg, D., Langdahl, B.L., Lips, P., Lorenc, R., Obermayer-Pietsch, B., Reeve, J., Reid, D.M., Amedei, A., Bassiti, A., Bustamante, M., Husted, L.B., Diez-Perez, A., Dobnig, H., Dunning, A.M., Enjuanes, A., Fahrleitner-Pammer, A., Fang, Y., Karczmarewicz, E., Kruk, M., Van Leeuwen, J.P., Mavilia, C., Van Meurs, J.B., Mangion, J., Mcguigan, F.E., Pols, H.A., Renner, W., Rivadeneira, F., Van Schoor, N.M., Scollen, S., Sherlock, R.E., Ioannidis, J.P., Investigators, A., Investigators, E., Investigators, E., Investigators, F., Investigators, L., Rotterdam Study, I., and Study, G. (2006). The association between common vitamin D receptor gene variations and osteoporosis: a participant-level meta-analysis. *Ann Intern Med* 145**,** 255-264.
547. Usenbo, A., Kramer, V., Young, T., and Musekiwa, A. (2015). Prevalence of Arthritis in Africa: A Systematic Review and Meta-Analysis. *PLoS One* 10**,** e0133858.
548. Van Den Berg, R., De Hooge, M., Rudwaleit, M., Sieper, J., Van Gaalen, F., Reijnierse, M., Landewe, R., Huizinga, T., and Van Der Heijde, D. (2013). ASAS modification of the Berlin algorithm for diagnosing axial spondyloarthritis: results from the SPondyloArthritis Caught Early (SPACE)-cohort and from the Assessment of SpondyloArthritis international Society (ASAS)-cohort. *Ann Rheum Dis* 72**,** 1646-1653.

549. Van Der Heijde, D., Baraf, H.S., Ramos-Remus, C., Calin, A., Weaver, A.L., Schiff, M., James, M., Markind, J.E., Reicin, A.S., Melian, A., and Dougados, M. (2005). Evaluation of the efficacy of etoricoxib in ankylosing spondylitis: results of a fifty-two-week, randomized, controlled study. *Arthritis Rheum* 52**,** 1205-1215.
550. Van Der Horst-Bruinsma, I.E., Clegg, D.O., and Dijkmans, B.A. (2002). Treatment of ankylosing spondylitis with disease modifying antirheumatic drugs. *Clin Exp Rheumatol* 20**,** S67-70.
551. Van Der Woude, D., Alemayehu, W.G., Verduijn, W., De Vries, R.R., Houwing-Duistermaat, J.J., Huizinga, T.W., and Toes, R.E. (2010). Gene-environment interaction influences the reactivity of autoantibodies to citrullinated antigens in rheumatoid arthritis. *Nat Genet* 42**,** 814-816; author reply 816.
552. Van Meurs, J.B., Trikalinos, T.A., Ralston, S.H., Balcells, S., Brandi, M.L., Brixen, K., Kiel, D.P., Langdahl, B.L., Lips, P., Ljunggren, O., Lorenc, R., Obermayer-Pietsch, B., Ohlsson, C., Pettersson, U., Reid, D.M., Rousseau, F., Scollen, S., Van Hul, W., Agueda, L., Akesson, K., Benevolenskaya, L.I., Ferrari, S.L., Hallmans, G., Hofman, A., Husted, L.B., Kruk, M., Kaptoge, S., Karasik, D., Karlsson, M.K., Lorentzon, M., Masi, L., Mcguigan, F.E., Mellstrom, D., Mosekilde, L., Nogues, X., Pols, H.A., Reeve, J., Renner, W., Rivadeneira, F., Van Schoor, N.M., Weber, K., Ioannidis, J.P., Uitterlinden, A.G., and Study, G. (2008). Large-scale analysis of association between LRP5 and LRP6 variants and osteoporosis. *JAMA* 299**,** 1277-1290.
553. Van Riel, P.L., and Renskers, L. (2016). The Disease Activity Score (DAS) and the Disease Activity Score using 28 joint counts (DAS28) in the management of rheumatoid arthritis. *Clin Exp Rheumatol* 34**,** S40-S44.
554. Vangay, P., Johnson, A.J., Ward, T.L., Al-Ghalith, G.A., Shields-Cutler, R.R., Hillmann, B.M., Lucas, S.K., Beura, L.K., Thompson, E.A., Till, L.M., Batres, R., Paw, B., Pergament, S.L., Saenyakul, P., Xiong, M., Kim, A.D., Kim, G., Masopust, D., Martens, E.C., Angkurawaranon, C., Mcgready, R., Kashyap, P.C., Culhane-Pera, K.A., and Knights, D. (2018). US Immigration Westernizes the Human Gut Microbiome. *Cell* 175**,** 962-972 e910.
555. Vasikaran, S., Eastell, R., Bruyere, O., Foldes, A.J., Garnero, P., Griesmacher, A., Mcclung, M., Morris, H.A., Silverman, S., Trenti, T., Wahl, D.A., Cooper, C., Kanis, J.A., and Group, I.-I.B.M.S.W. (2011). Markers of bone turnover for the prediction of fracture risk and monitoring of osteoporosis treatment: a need for international reference standards. *Osteoporos Int* 22**,** 391-420.
556. Veale, D., Yanni, G., Rogers, S., Barnes, L., Bresnihan, B., and Fitzgerald, O. (1993). Reduced synovial membrane macrophage numbers, ELAM-1 expression, and lining layer hyperplasia in psoriatic arthritis as compared with rheumatoid arthritis. *Arthritis Rheum* 36**,** 893-900.
557. Veale, D.J., and Fearon, U. (2018). The pathogenesis of psoriatic arthritis. *Lancet* 391**,** 2273-2284.
558. Veale, D.J., Ritchlin, C., and Fitzgerald, O. (2005). Immunopathology of psoriasis and psoriatic arthritis. *Ann Rheum Dis* 64 Suppl 2**,** ii26-29.
559. Venken, K., Hellings, N., Thewissen, M., Somers, V., Hensen, K., Rummens, J.L., Medaer, R., Hupperts, R., and Stinissen, P. (2008). Compromised CD4+ CD25(high) regulatory T-cell function in patients with relapsing-remitting multiple sclerosis is correlated with a reduced frequency of FOXP3-positive

cells and reduced FOXP3 expression at the single-cell level. *Immunology* 123**,** 79-89.

560. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., Mckusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Francesco, V.D., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.-R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z.Y., Wang, A., Wang, X., Wang, J., Wei, M.-H., Wides, R., Xiao, C., Yan, C., et al. (2001). The Sequence of the Human Genome. *Science* 291**,** 1304-1351.
561. Vescovi, P., and Nammour, S. (2010). Bisphosphonate-Related Osteonecrosis of the Jaw (BRONJ) therapy. A critical review. *Minerva Stomatol* 59**,** 181-203, 204-113.
562. Viatte, S., Massey, J., Bowes, J., Duffus, K., Arc, O.C., Eyre, S., Barton, A., and Worthington, J. (2016). Replication of Associations of Genetic Loci Outside the HLA Region With Susceptibility to Anti-Cyclic Citrullinated Peptide-Negative Rheumatoid Arthritis. *Arthritis Rheumatol* 68**,** 1603-1613.
563. Videm, V., Cortes, A., Thomas, R., and Brown, M.A. (2014). Current smoking is associated with incident ankylosing spondylitis -- the HUNT population-based Norwegian health study. *J Rheumatol* 41**,** 2041-2048.
564. Viglietta, V., Baecher-Allan, C., Weiner, H.L., and Hafler, D.A. (2004). Loss of functional suppression by CD4+CD25+ regulatory T cells in patients with multiple sclerosis. *J Exp Med* 199**,** 971-979.
565. Viladomiu, M., Kivolowitz, C., Abdulhamid, A., Dogan, B., Victorio, D., Castellanos, J.G., Woo, V., Teng, F., Tran, N.L., Sczesnak, A., Chai, C., Kim, M., Diehl, G.E., Ajami, N.J., Petrosino, J.F., Zhou, X.K., Schwartzman, S., Mandl, L.A., Abramowitz, M., Jacob, V., Bosworth, B., Steinlauf, A., Scherl, E.J., Wu, H.J., Simpson, K.W., and Longman, R.S. (2017). IgA-coated E. coli enriched in Crohn's disease spondyloarthritis promote TH17-dependent inflammation. *Sci Transl Med* 9.
566. Vince, A., Dyer, N.H., O'grady, F.W., and Dawson, A.M. (1972). Bacteriological studies in Crohn's disease. *J Med Microbiol* 5**,** 219-229.
567. Vinolo, M.A., Rodrigues, H.G., Nachbar, R.T., and Curi, R. (2011). Regulation of inflammation by short chain fatty acids. *Nutrients* 3**,** 858-876.
568. Vojdani, A., Pollard, K.M., and Campbell, A.W. (2014). Environmental triggers and autoimmunity. *Autoimmune Dis* 2014**,** 798029.
569. Wade, S.W., Strader, C., Fitzpatrick, L.A., Anthony, M.S., and O'malley, C.D. (2014). Estimating prevalence of osteoporosis: examples from industrialized countries. *Arch Osteoporos* 9**,** 182.

570. Wahlstrom, A., Sayin, S.I., Marschall, H.U., and Backhed, F. (2016). Intestinal Crosstalk between Bile Acids and Microbiota and Its Impact on Host Metabolism. *Cell Metab* 24**,** 41-50.
571. Waljee, A.K., Wiitala, W.L., Govani, S., Stidham, R., Saini, S., Hou, J., Feagins, L.A., Khan, N., Good, C.B., Vijan, S., and Higgins, P.D. (2016). Corticosteroid Use and Complications in a US Inflammatory Bowel Disease Cohort. *PLoS One* 11**,** e0158017.
572. Wallace, J.L., Syer, S., Denou, E., De Palma, G., Vong, L., Mcknight, W., Jury, J., Bolla, M., Bercik, P., Collins, S.M., Verdu, E., and Ongini, E. (2011). Proton pump inhibitors exacerbate NSAID-induced small intestinal injury by inducing dysbiosis. *Gastroenterology* 141**,** 1314-1322, 1322 e1311-1315.
573. Wallin, M.T., Culpepper, W.J., Campbell, J.D., Nelson, L.M., Langer-Gould, A., Marrie, R.A., Cutter, G.R., Kaye, W.E., Wagner, L., Tremlett, H., Buka, S.L., Dilokthornsakul, P., Topol, B., Chen, L.H., Larocca, N.G., and Workgroup, U.S.M.S.P. (2019). The prevalence of MS in the United States: A population-based estimate using health claims data. *Neurology* 92**,** e1029-e1040.
574. Wanders, A., Heijde, D., Landewe, R., Behier, J.M., Calin, A., Olivieri, I., Zeidler, H., and Dougados, M. (2005). Nonsteroidal antiinflammatory drugs reduce radiographic progression in patients with ankylosing spondylitis: a randomized clinical trial. *Arthritis Rheum* 52**,** 1756-1765.
575. Wang, J., Wang, Y., Gao, W., Wang, B., Zhao, H., Zeng, Y., Ji, Y., and Hao, D. (2017). Diversity analysis of gut microbiota in osteoporosis and osteopenia patients. *PeerJ* 5**,** e3450.
576. Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73**,** 5261-5267.
577. Wang, Z., Friedrich, C., Hagemann, S.C., Korte, W.H., Goharani, N., Cording, S., Eberl, G., Sparwasser, T., and Lochner, M. (2014). Regulatory T cells promote a protective Th17-associated immune response to intestinal bacterial infection with C. rodentium. *Mucosal Immunol* 7**,** 1290-1301.
578. Warner, B.B. (2019). The contribution of the gut microbiome to neurodevelopment and neuropsychiatric disorders. *Pediatr Res* 85**,** 216-224.
579. Watanabe, T., Kitani, A., Murray, P.J., and Strober, W. (2004). NOD2 is a negative regulator of Toll-like receptor 2-mediated T helper type 1 responses. *Nat Immunol* 5**,** 800-808.
580. Watanabe, Y., Nagai, F., and Morotomi, M. (2012). Characterization of Phascolarctobacterium succinatutens sp. nov., an asaccharolytic, succinate-utilizing bacterium isolated from human feces. *Appl Environ Microbiol* 78**,** 511-518.
581. Wehkamp, J., Harder, J., Weichenthal, M., Mueller, O., Herrlinger, K.R., Fellermann, K., Schroeder, J.M., and Stange, E.F. (2003). Inducible and constitutive beta-defensins are differentially expressed in Crohn's disease and ulcerative colitis. *Inflamm Bowel Dis* 9**,** 215-223.
582. Weinstein, L. (1961). Bacteriologic aspects of ulcerative colitis. *Gastroenterology* 40**,** 323-330.
583. Weiss, S., Xu, Z.Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J.R., Vazquez-Baeza, Y., Birmingham, A., Hyde, E.R.,

and Knight, R. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 5**,** 27.
584. Weitzmann, M.N., and Pacifici, R. (2006). Estrogen deficiency and bone loss: an inflammatory tale. *J Clin Invest* 116**,** 1186-1194.
585. Whibley, N., Tucci, A., and Powrie, F. (2019). Regulatory T cell adaptation in the intestine and skin. *Nat Immunol* 20**,** 386-396.
586. Whittaker, R.H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs* 30**,** 279-338.
587. Whittaker, R.H. (1972). Evolution and Measurement of Species Diversity. *Taxon* 21**,** 213-251.
588. Wilkins, T., Jarvis, K., and Patel, J. (2011). Diagnosis and management of Crohn's disease. *Am Fam Physician* 84**,** 1365-1375.
589. Willemsen, L.E., Koetsier, M.A., Van Deventer, S.J., and Van Tol, E.A. (2003). Short chain fatty acids stimulate epithelial mucin 2 expression through differential effects on prostaglandin E(1) and E(2) production by intestinal myofibroblasts. *Gut* 52**,** 1442-1447.
590. Wiren, K.M., Zhang, X.W., Olson, D.A., Turner, R.T., and Iwaniec, U.T. (2012). Androgen prevents hypogonadal bone loss via inhibition of resorption mediated by mature osteoblasts/osteocytes. *Bone* 51**,** 835-846.
591. Woese, C.R., and Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences* 74**,** 5088-5090.
592. Woese, C.R., Kandler, O., and Wheelis, M.L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* 87**,** 4576-4579.
593. Wood, D.E., and Salzberg, S.L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 15**,** R46.
594. Wostmann, B.S. (1981). The germfree animal in nutritional studies. *Annu Rev Nutr* 1**,** 257-279.
595. Wu, H.J., Ivanov, Ii, Darce, J., Hattori, K., Shima, T., Umesaki, Y., Littman, D.R., Benoist, C., and Mathis, D. (2010). Gut-residing segmented filamentous bacteria drive autoimmune arthritis via T helper 17 cells. *Immunity* 32**,** 815-827.
596. Wu, S., Rhee, K.J., Albesiano, E., Rabizadeh, S., Wu, X., Yen, H.R., Huso, D.L., Brancati, F.L., Wick, E., Mcallister, F., Housseau, F., Pardoll, D.M., and Sears, C.L. (2009). A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nat Med* 15**,** 1016-1022.
597. Wu, Y.W., Simmons, B.A., and Singer, S.W. (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32**,** 605-607.
598. Wu, Y.W., Tang, Y.H., Tringe, S.G., Simmons, B.A., and Singer, S.W. (2014). MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* 2**,** 26.
599. Xavier, R.J., and Podolsky, D.K. (2007). Unravelling the pathogenesis of inflammatory bowel disease. *Nature* 448**,** 427-434.
600. Xu, X., Jia, X., Mo, L., Liu, C., Zheng, L., Yuan, Q., and Zhou, X. (2017). Intestinal microbiota: a potential target for the treatment of postmenopausal osteoporosis. *Bone Res* 5**,** 17046.

601. Yang, B., Wang, Y., and Qian, P.Y. (2016a). Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics* 17**,** 135.
602. Yang, L., Wang, L., Wang, X., Xian, C.J., and Lu, H. (2016b). A Possible Role of Intestinal Microbiota in the Pathogenesis of Ankylosing Spondylitis. *Int J Mol Sci* 17.
603. Yarza, P., Yilmaz, P., Pruesse, E., Glockner, F.O., Ludwig, W., Schleifer, K.H., Whitman, W.B., Euzeby, J., Amann, R., and Rossello-Mora, R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* 12**,** 635-645.
604. Ye, Y., and Doak, T.G. (2009). A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS Comput Biol* 5**,** e1000465.
605. Yokote, H., Miyake, S., Croxford, J.L., Oki, S., Mizusawa, H., and Yamamura, T. (2008). NKT cell-dependent amelioration of a mouse model of multiple sclerosis by altering gut flora. *Am J Pathol* 173**,** 1714-1723.
606. Yoon, H.M., Suh, C.H., Kim, J.R., Lee, J.S., Jung, A.Y., Kim, K.M., and Cho, Y.A. (2017). Diagnostic Performance of Magnetic Resonance Enterography for Detection of Active Inflammation in Children and Adolescents With Inflammatory Bowel Disease: A Systematic Review and Diagnostic Meta-analysis. *JAMA Pediatr* 171**,** 1208-1216.
607. Yoshioka, H., Iseki, K., and Fujita, K. (1983). Development and differences of intestinal flora in the neonatal period in breast-fed and bottle-fed infants. *Pediatrics* 72**,** 317-321.
608. Yu, F., Sharma, S., Edwards, J., Feigenbaum, L., and Zhu, J. (2015). Dynamic expression of transcription factors T-bet and GATA-3 by regulatory T cells maintains immunotolerance. *Nat Immunol* 16**,** 197-206.
609. Yu, L.C. (2018). Microbiota dysbiosis and barrier dysfunction in inflammatory bowel disease and colorectal cancers: exploring a common ground hypothesis. *J Biomed Sci* 25**,** 79.
610. Zamani, B., Golkar, H.R., Farshbaf, S., Emadi-Baygi, M., Tajabadi-Ebrahimi, M., Jafari, P., Akhavan, R., Taghizadeh, M., Memarzadeh, M.R., and Asemi, Z. (2016). Clinical and metabolic response to probiotic supplementation in patients with rheumatoid arthritis: a randomized, double-blind, placebo-controlled trial. *Int J Rheum Dis* 19**,** 869-879.
611. Zeng, Q., Junli, G., Liu, X., Chen, C., Sun, X., Li, H., Zhou, Y., Cui, C., Wang, Y., Yang, Y., Wu, A., Shu, Y., Hu, X., Lu, Z., Zheng, S.G., Qiu, W., and Lu, Y. (2019). Gut dysbiosis and lack of short chain fatty acids in a Chinese cohort of patients with multiple sclerosis. *Neurochem Int* 129**,** 104468.
612. Zerbino, D.R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18**,** 821-829.
613. Zhang, C., Li, L., Feng, K., Fan, D., Xue, W., and Lu, J. (2017). 'Repair' Treg Cells in Tissue Injury. *Cell Physiol Biochem* 43**,** 2155-2169.
614. Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30**,** 614-620.
615. Zhang, S. (2018). The role of transforming growth factor beta in T helper 17 differentiation. *Immunology* 155**,** 24-35.
616. Zhang, X., Zhang, D., Jia, H., Feng, Q., Wang, D., Liang, D., Wu, X., Li, J., Tang, L., Li, Y., Lan, Z., Chen, B., Li, Y., Zhong, H., Xie, H., Jie, Z., Chen, W.,

Tang, S., Xu, X., Wang, X., Cai, X., Liu, S., Xia, Y., Li, J., Qiao, X., Al-Aama, J.Y., Chen, H., Wang, L., Wu, Q.J., Zhang, F., Zheng, W., Li, Y., Zhang, M., Luo, G., Xue, W., Xiao, L., Li, J., Chen, W., Xu, X., Yin, Y., Yang, H., Wang, J., Kristiansen, K., Liu, L., Li, T., Huang, Q., Li, Y., and Wang, J. (2015). The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nat Med* 21**,** 895-905.

617. Zhang, Y.Z., and Li, Y.Y. (2014). Inflammatory bowel disease: pathogenesis. *World J Gastroenterol* 20**,** 91-99.
618. Zhernakova, A., Kurilshikov, A., Bonder, M.J., Tigchelaar, E.F., Schirmer, M., Vatanen, T., Mujagic, Z., Vila, A.V., Falony, G., Vieira-Silva, S., Wang, J., Imhann, F., Brandsma, E., Jankipersadsing, S.A., Joossens, M., Cenit, M.C., Deelen, P., Swertz, M.A., Lifelines Cohort, S., Weersma, R.K., Feskens, E.J., Netea, M.G., Gevers, D., Jonkers, D., Franke, L., Aulchenko, Y.S., Huttenhower, C., Raes, J., Hofker, M.H., Xavier, R.J., Wijmenga, C., and Fu, J. (2016). Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* 352**,** 565-569.
619. Zhou, D., Zhang, H., Bai, Z., Zhang, A., Bai, F., Luo, X., Hou, Y., Ding, X., Sun, B., Sun, X., Ma, N., Wang, C., Dai, X., and Lu, Z. (2016). Exposure to soil, house dust and decaying plants increases gut microbial diversity and decreases serum immunoglobulin E levels in BALB/c mice. *Environ Microbiol* 18**,** 1326-1337.
620. Zhou, L., Ivanov, Ii, Spolski, R., Min, R., Shenderov, K., Egawa, T., Levy, D.E., Leonard, W.J., and Littman, D.R. (2007). IL-6 programs T(H)-17 cell differentiation by promoting sequential engagement of the IL-21 and IL-23 pathways. *Nat Immunol* 8**,** 967-974.
621. Zhu, W., He, X., Cheng, K., Zhang, L., Chen, D., Wang, X., Qiu, G., Cao, X., and Weng, X. (2019). Ankylosing spondylitis: etiology, pathogenesis, and treatments. *Bone Research* 7**,** 22.
622. Zhu, W., Lomsadze, A., and Borodovsky, M. (2010). Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res* 38**,** e132.
623. Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R., and Goodman, A.L. (2019). Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature* 570**,** 462-467.
624. Zmora, N., Suez, J., and Elinav, E. (2019). You are what you eat: diet, health and the gut microbiota. *Nat Rev Gastroenterol Hepatol* 16**,** 35-56.

Chapter II

Gut Microbiota Alterations associated with reduced Bone Mineral Density in Older Adults

This chapter has been published as

Mrinmoy Das, Owen Cronin, David M. Keohane, Edel M. Cormac, Helena Nugent, Michelle Nugent, Catherine Molloy, Paul W. O'Toole, Fergus Shanahan, Michael G. Molloy, Ian B. Jeffery

Gut Microbiota Alterations associated with reduced Bone Mineral Density in Older Adults

August 2019, Rheumatology (Oxford Academic)

PMID: 31378815, https://doi.org/10.1093/rheumatology/kez302

Contents

Chapter 2

Chapter 2 Gut Microbiota Alterations associated with reduced Bone Mineral Density in Older Adults

2.1 Abstract

Objective To investigate compositional differences in the gut microbiota associated with bone homeostasis and fractures in a cohort of older adults.

Methods Faecal microbiota profiles were determined from 181 individuals with osteopenia (n=61), osteoporosis (n=60) and an age- and gender-matched group with normal bone mineral density (n=60). Analysis of the 16S (V3-V4 region) amplicon dataset classified to the Genus level was used to identify significantly differentially abundant taxa. Adjustments were made for potential confounding variables identified from the literature using several statistical models.

Results We identified six genera that were significantly altered in abundance in osteoporosis or osteopenic goups compared to age- and gender-matched controls. A detailed study of microbiota associations with meta-data variables which included BMI, health status, diet and medication, revealed that these meta-data explained 15-19% of the variance within the microbiota dataset. Bone mineral density measurements were significantly associated with alterations in the microbiota. After controlling for known biological confounders, five of the six taxa remained significant. Overall microbiota alpha diversity did not correlate to bone mineral density in this study.

Conclusion Reduced bone mineral density in osteopenia and osteoporosis is associated with an altered microbiota. These alterations may be useful as biomarkers or therapeutic targets in individuals at high risk of reductions in bone mineral density. These observations will lead to a better understanding of the relationship between the microbiota and bone homeostasis.

Keywords

Osteoporosis, Gut microbiota**,** Bone mineral density, Elderly, Osteopenia

Key messages

Reduced bone mineral density is associated with taxon-specific signatures in the gut microbiota.

Medication, anthropometric measures, nutrition, and gender are associated with gut microbiota composition.

Confounders do not explain the microbiota-bone density interactions observed here.

2.2 Introduction

Osteoporosis, characterized by reduced bone mineral density and degradation of the micro-architectural structure of bone, affects over 27.5 million people in Europe (Hernlund et al., 2013). Over the age of 50, 1 in 3 women and 1 in 5 men, worldwide, will experience an osteoporotic fracture in their lifetime, representing a significant burden for patients and health care providers (Sozen et al., 2017). The aetiology of osteoporosis and its precursor, osteopenia, is multi-factorial. Contributing factors include oestrogen and vitamin D deficiency, and genetic modification in regulatory genes such as vitamin D receptors and transforming growth factor-β (Eastell et al., 2016). Osteoporosis occurrence is accelerated in patients with immune-mediated inflammatory conditions where excessive production of pro-inflammatory cytokines leads to increased osteoclastic bone resorption (e.g. inflammatory bowel disease, rheumatoid arthritis, and ankylosing spondylitis) (Donnelly et al., 1994;Sambrook et al., 1995;Ali et al., 2009). The gut microbiome is known to modulate immune cell activities and alterations in the microbiome have previously been associated with these inflammatory conditions (Clemente et al., 2018).

The gut microbiome shares a complex relationship with the host. Development and maturation of the innate and adaptive immunity in the host is dependent on appropriate exposure to the gut microbiota (Peterson et al., 2015). Alterations in the microbiota may result in immune system modulation or activation. Circulating osteoclastogenic cytokines may be increased in a T cell-dependent mechanism by the microbiota which can drive bone resorption in inflammatory conditions (Hsu and Pacifici, 2018). Several investigations have identified microbes that regulate the production of hormones or improve uptake of vitamins that are integral to bone health (Jones et al., 2013;Baker et al., 2017).

Studies with germ-free and antibiotic-treated animals have indicated the possibility of gut microbial influence on both bone mass accumulation and turnover. These animals have shown a reduction in osteoclastic precursor cell number (Sjogren et al., 2012), an increase in bone mass (Nobel et al., 2015) and improvement in bone strength and material properties (Pytlik et al., 2004).

We have previously identified significant microbiota alterations associated with inflamm-aging and frailty in an elderly cohort (Claesson et al., 2012). Other studies have demonstrated that the absence of gut microbiota leads to a reduction in bone mechanical strength (Guss et al., 2017) and inversely, long term colonisation of pathogen-free gut microbiota increases bone formation (Yan et al., 2016). In contrast another recent study suggested that microbiota restoration in germ-free mice does not affect bone loss (Quach et al., 2018). These conflicting findings may, in part, be due to different animal genotypes, the anti-microbials administered and the absence or presence of particular taxa in their baseline microbiota.

Our aim in the present study was to determine whether gut microbiota features are associated with bone mineral density in a cohort of individuals at high risk of reduced bone mineral density and fractures. In addition to this, any genus-level taxa associated with altered bone mineral density would be identified by comparing the gut microbiota composition of osteopenic and osteoporotic patients with those of age- and gender-matched controls with normal bone mineral density. Our hypothesis was that intestinal microbiota composition was different in the osteoporotic subjects. Furthermore, we developed and applied a rigorous statistical regime to remove the effect of potentially confounding variables.

2.3 Methods

2.3.1 Subject recruitments and clinical information

Ethical approval was granted by the Clinical Research Ethics Committee of the Cork Teaching Hospitals (CREC) before recruitment. Adult female and male subjects, aged 55 to 75 years, were recruited from the bone densitometry unit at Cork University Hospital, Cork City, Ireland. The indications for referral for bone mineral density (BMD) assessment by Dual-energy X-ray absorptiometry (DEXA) were varied with referrals from primary, secondary and tertiary care. No single specific referral criterion was used, and request for assessment was at the discretion of the attending clinician and not the study investigators. Written informed consent was obtained from the participants. Individuals with a known history of alcohol abuse, participation in an investigational drug trial in the 30 days before enrolment, use of antibiotics in the 3 months prior to bone density measurement, and previous partial or total colectomy were excluded. No measure was taken to exclude participants with co-existing osteoarthritis, aortic calcification or fractures. Altogether from 193 participants, stool samples were collected. Due to lack of vitamin D information from 12 samples, they were excluded from the analysis, resulting in the final dataset comprising of 181 participants.

Patients underwent DEXA assessment of Bone mineral density (BMD) (g/cm^2) at the femoral neck, and antero-posterior lumbar spine (L1-L4) with a GE Healthcare Lunar iDXA machine (GE Healthcare, Madison, WI) and enCORE software (V.13.4, 2010) using standardized methodology (Hind et al., 2010). T-score threshold was used to define 3 groups based on their BMD. These were normal BMD (n = 60) with a T-score of ≥ -1, patients with osteopenia (n = 61) with a T-score between -1 and -2.5, and patients with osteoporosis (n = 60) were defined as having a T-score of ≤ -2.5

(Kanis et al., 1994;Kanis et al., 2013). The detailed procedure of recording anthropometric, clinical, dietary and medications information is recorded in the supplementary text.

2.3.2 Molecular methods and bioinformatics

Genomic DNA was extracted from 0.25g of each of the faecal samples based on a modified Yu and Morrison protocol (Yu and Morrison, 2004). The V3-V4 region of the 16S rRNA gene was amplified and sequenced (Klindworth et al., 2013) on the Illumina MiSeq platform at Moorepark Teagasc Food Research Centre, Fermoy, Ireland. The reads were merged using flash (v1.2.8) (Magoc and Salzberg, 2011). The forward adapters were removed using cutadapt (v1.8.3). The quality filtering of reads and removal of reverse primers were carried out using the QIIME (v1.9.1) (Caporaso et al., 2010) pipeline with default settings. The removal of chimeric sequences and generation of operational taxonomic units (OTUs) at 97% identity threshold was done using USEARCH (v8.1) (Edgar, 2010). Representative OTUs were classified using the Ribosomal Database Project (RDP) database (v11.4) (Cole et al., 2014) implemented in mothur (v1.34.4) (Schloss et al., 2009). α- and β-diversity measures were produced from a rarefied dataset (10,613 reads per sample).

2.3.3 Statistical analysis

All statistical analyses were carried out in the R statistical software (v3.4.0) (Team, 2017). Significance was determined by a cut-off p-value ≤ 0.05 and p-adjusted ≤ 0.05 (Benjamini-Hochberg procedure) unless stated otherwise. P-adjusted for pairwise comparison is based on the p-values obtained from all the pairwise comparisons for each variable.

2.3.4 Analysis of meta-data

Kruskal-Wallis, Dunn's test (v1.3.4) (Dinno, 2017) and/or chi-square tests were carried out to identify anthropometric, clinical, dietary and medications significantly different between the groups. For chi-square testing, at least 7 participants were present across the whole dataset for that factor.

2.3.5 Analysis of microbiota data

Kruskal-Wallis test was used to determine significant difference in α-diversity measures between the groups. Co-inertia analysis was used to explore the covariance between the dietary dataset and microbiota dataset. DESeq2 (v1.16.1) (Love et al., 2014) was used to identify differentially abundant taxa from the microbiota dataset. The dataset was filtered to retain only those taxa which were present in at least 20% of the samples across the whole dataset. A DESeq2 model adjusted for Body Mass Index (BMI) and gender was used to identify genera that were significantly differentially abundant.

2.3.6 Identification of meta-data variables associated with beta-diversity

Meta-data variables significantly associated with variations in global microbiota profiles were identified using permutational multivariate analysis of variance (PERMANOVA). A nominal p-value of ≤ 0.05 was used as the analysis was a confirmation of previously established associations. Subjects with diseases such as coeliac disease, diverticulitis and inflammatory arthritis conditions were present within the dataset and tested separately. Inflammatory and non-inflammatory diseases can alter the microbiota with a common dysbiosis signature (Duvallet et al., 2017). To

investigate the common signature of microbiota-associated inflammatory diseases, we created an inflammatory disease index, where the presence of any one of the microbiota-associated conditions (Coeliac, Diverticulitis, Arthritis, IBD and Multiple Sclerosis) was considered. Nominally significant meta-data variables were added to a single PERMANOVA model to identify overall effect sizes. The cumulative effect was calculated based on these pre-defined groups of variables.

2.3.7 Analysis of confounding variables

Clinical variables which have been reported to interact with the microbiota were identified from the literature (**Supplementary table 1**). These included diet (Healthy food diversity (HFD) index) (Drescher et al., 2007), Barthel score (Mahoney and Barthel, 1965), Godin leisure time activity score (Godin and Shephard, 1985), Mini-Mental State Examination (MMSE) scores (Molloy et al., 1991), Mini Nutritional Assessment (MNA) (Guigoz and Vellas, 1999) and Carlson co-morbidity index (Charlson et al., 1987). Secondly, the meta-data identified as significantly different between the subject groups were confirmed by a literature search (**Table 1, Supplementary table 2**, p-adjusted ≤ 0.05) and were added to the analysis as potential confounders.

Confounding factors were modelled using a general linear mixed-effect model (GLMM) using the negative binomial distribution and the sequencing depth was controlled for by categorising the number of reads into four quartiles and adding this information as a random effect to the model. Firstly, univariate GLMMs were generated with individual confounding factors as the predictor and the significant taxa as the response. The confounders identified as significant for individual taxa were controlled for in a bivariate model. To maximise the number of known confounders

identified, a nominal p-value was regarded as significant. In this model, the effect of group category was evaluated after adjusted for the individual significant confounders. Summary reports were generated for both the univariate and bivariate GLMM models to explain the contribution of the predictors.

An expanded methodology is available in the supplementary notes.

2.4 Results

2.4.1 Descriptive statistics of the study population

In the present study, samples and clinical information for 181 individuals were analysed. These patients were evenly divided between those with normal BMD (60), osteopenia (61) and osteoporosis (60) groups. Clinical, physiological, biomedical and dietary measures were investigated and significant differences between normal BMD, osteopenia and osteoporosis participants were detected. Differences in bone density measurements (T-score and BMD of the anterior-posterior spine and neck of femur) were confirmed and differences in BMI, weight, circumference measures, vitamin D levels, and the use of calcium and bisphosphonate supplements were noted **(Table 1, Supplementary Table 2, Supplementary Figure 1, 2A-B)**. Due to the recruitment by clinical referral of this high risk cohort, there was a high rate of fractures in all groupings with percentages for one or more fractures being 40% (24/60), 59% (36/61) and 42% (25/60) for normal BMD, osteopenia and osteoporosis groups respectively and percentages for 2 or more fractures being 7% (4/60), 23% (14/61) and 15% (9/60) respectively.

Table 1 Significant characteristics of the participants in the final dataset

Meta-data	Healthy (n=60)	Osteopenia (n=61)	Osteoporosis (n=60)	Significance
Gender (Male/Female)	**13/47**	**7/54**	**11/49**	NS
Age (Years)	63.57±5.73	64.84±5.28	65.07±5.58	NS
BMI	29.09±4.57	27.20±4.80	23.96±3.31	***
Weight (kg)	78.86±13.60	70.96±14.44	61.65±9.44	***
Waist circumference (cm)	95.71±11.95 (13/46)	89.81±12.40 (6/54)	81.81±9.36	***
Hip circumference (cm)	106.71±9.83 (13/46)	103.63±10.45 (6/53)	96.66±7.26	***
Waist-Hip ratio	0.90±0.08 (13/46)	0.87±0.06 (6/53)	0.85±0.07	**
Mid arm circumference (cm)	30.98±3.62 (12/47)	28.85±3.97	26.80±2.91	***
Calf circumference (cm)	37.69±3.73 (11/47)	35.76±4.28	33.93±2.76	***
AP spine T-score	0.28±1.02	-1.16±0.87	-2.86±0.74	***
AP spine BMD (g/cm^2)	1.22±0.13	1.04±0.11	0.84±0.09	***
Neck-femur T-score	-0.54±0.35	-1.27±0.53	-1.95±0.80	***
Neck-femur BMD (g/cm^2)	0.98±0.09	0.84±0.07	0.84±0.68	***
Vitamin D3 [25(OH)D3] (nmol/L)	60.49±20.84	69.98±25.27	75.96±26.43	**
Total Vitamin D [25(OH)D)] (nmol/L)	63.68±20.57	72.40±25.36	79.18±26.07	**
Calcium supplements (Yes/No)	10/50	31/30	35/25	***
Bisphosphonate medication (Yes/No)	4/56	6/55	17/43	***

Group-wise comparisons of the clinical variables. Kruskal-Wallis or chi-square statistic was used to determine significance. The values represent mean and standard deviation: Mean±SD or number of samples per group.

BMI, body mass index; BMD, bone mineral density; 25(OH)D3, vitamin D3, Total 25(OH)D, total vitamin D.

Significance: p-adjusted *** ≤ 0.0005, ** ≤ 0.005, NS Not significant

Values in brackets for circumference measures and waist-hip ratio represents different sample size. The complete list of sample characteristics along with pairwise comparisons is available in **supplementary table 2**.

Indicators of BMD measures: AP spine T-score, AP spine BMD (g/cm2), Neck-femur T-score, Neck-femur BMD (g/cm2)

Indicators of obesity: BMI, Weight (kg), Waist circumference (cm), Hip circumference (cm), Waist-Hip ratio, Mid arm circumference (cm), Calf circumference (cm)

2.4.2 Microbiota characterisation

The microbiota composition of the samples analysed was dominated by phylum Firmicutes with a mean abundance of 78.9% across the whole dataset followed in rank abundance order by Bacteroidetes accounting for 14.9%. Other phyla accounted for 5.8%, while 0.4% were unclassified **(Supplementary figure 3A)**. The core microbiota consisted of 23 genera that were found in at least 90% of the samples. The top five genera with mean relative abundance in the whole dataset were *Faecalibacterium* (11.7%), *Bacteroides* (9.4%), *Roseburia* (7.9%), *Blautia* (7.6%), and *Coprococcus* (3.2%) **(Supplementary figure 3B)**. Based on PCoA (Principal Coordinate analysis) on different β-diversity measures, Axes 1 and 2 explained 11-17% and 8-13% of variance respectively (**Supplementary figure 4A, Supplementary Table 3**). The relationship of BMD measures with global microbiota profile was visualised using dbRDA (distance-based redundancy analysis), testing AP spine BMD measure with Bray-Curtis distance (**Supplementary figure 4B**). With regard to α-diversity, an average richness of 308.7±84.2 was observed and extrapolated richness (chao1) was estimated at 406.8±122 (**Figure 1D, Supplementary figure 4C**) No significant difference was observed in any of the alpha diversity indices among the three clinical groups (**Figure 1D-E, Supplementary Figure 4C-D**).

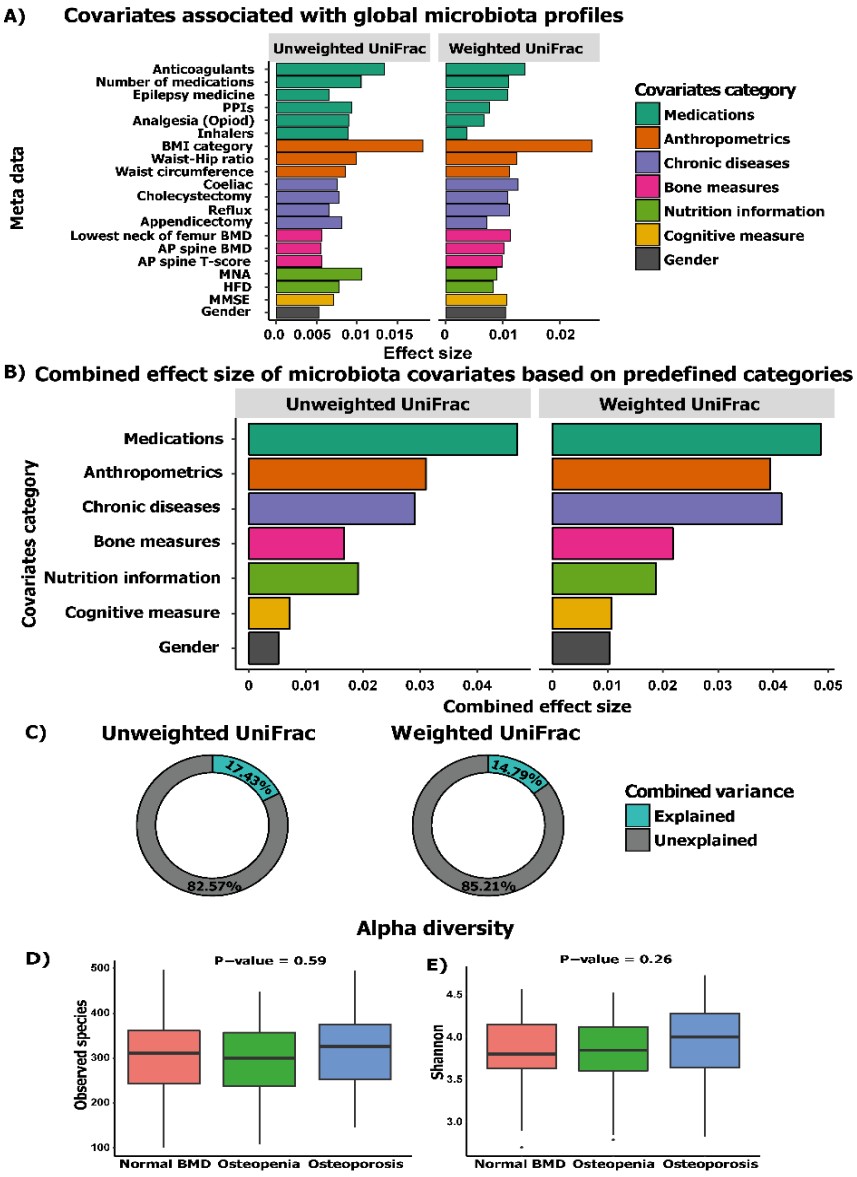


Figure 1 Effect size of covariates significantly associated with global microbiota profiles

Significance was defined as a p-value of 0.05. **A)** A total of 20 factors were identified to be nominally significantly associated with β–diversity. The bar plot shows the variation explained by each factor individually on microbiota composition (Weighted and Unweighted UniFrac). The factors are sorted based on their mean cumulative (grouped into predefined categories) and individual effect size from both distance measures. **B)** shows the combined variance explained by the predefined categories. **C)** The donut plot shows the portion of combined variance explained by the nominally significant factors on weighted and unweighted UniFrac measures respectively. **D)** and **E)** shows the lack of significant difference in observed species diversity measure and Shannon index respectively.

PPIs, Proton Pump Inhibitors; BMI, Body Mass Index; BMD, Bone mineral density; MNA, Mini Nutritional Assessment; HFD, Healthy Food Diversity; MMSE, Mini-Mental State Examination

2.4.3 Association of gut microbiota with covariates

Both sets of bone density measurements and one of the T-scores tested explained a significant amount of microbiota variance (p-value ≤ 0.05), verifying the original hypothesis that BMD is associated with alterations in the microbiota **(Figure 1, Supplementary Table 4)**. We extended this beta-diversity analysis to known microbiota-associated putative meta-data variables to measure their effect on the microbiota **(Supplementary table 4)**. This analysis identified 20 meta-data variables to be associated with the global microbiota profile **(Figure 1A)**, with BMI having the largest effect size individually (2.1%). An inflammatory disease index was created indicating the presence or absence of a disease, disorder or condition. This index showed a significant association with the β-diversity (Bray-Curtis p-value: 0.042, R2: 0.009).

Amongst the significant variables, the combined effect-size of the different medications explained the most variance (4.8%) followed by anthropometric measures (3.5%). Chronic diseases explained 3.5% and BMD measurement was the fourth largest contributor to effect-size (2%) Nutritional information (HFD and MNA), cognitive measures (MMSE) and gender explained 1.4%, 1% and 0.6% of variance respectively **(Figure 1B)**. Overall, a cumulative total range of 15-17% of the variance in our dataset was explained which indicates that stochastic factors explain the majority of the variance in global microbiota composition **(Figure 1C)**. Analysis of the Food Frequency Questionnaire (FFQ) data and diet quality as measured by the HFD index revealed no significant difference in diet composition or HFD across the three groups. Co-inertia analysis of the FFQ dataset with the microbiota dataset (**Figure 2A**) graphically confirmed a significant co-variation between the two datasets which was independent of the defined bone health groups.

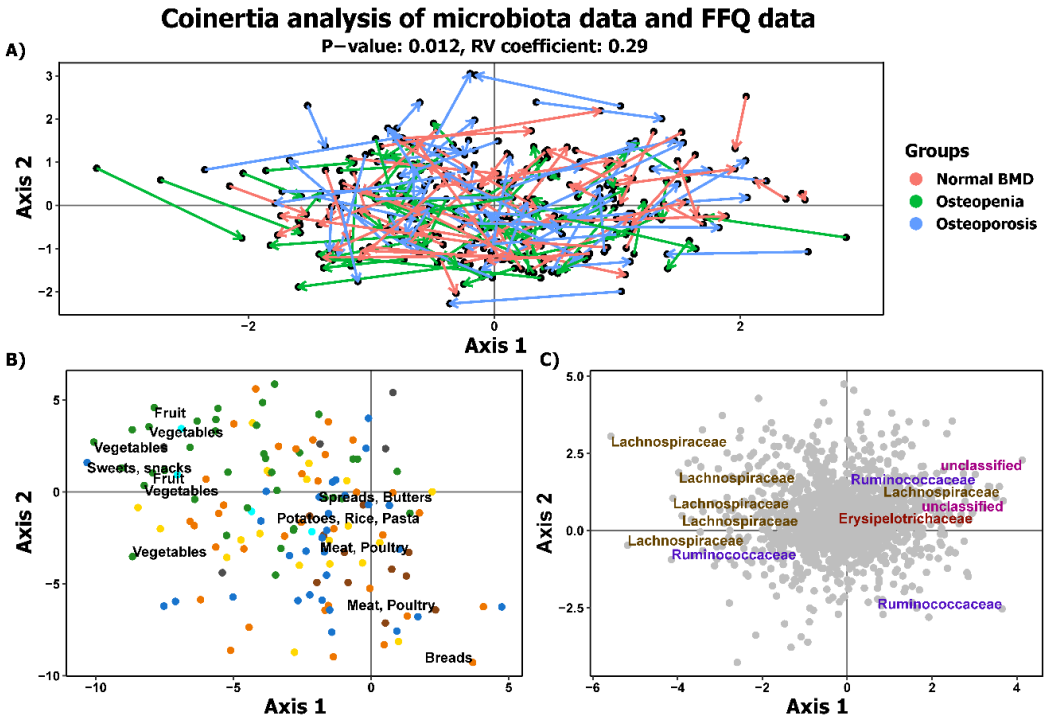


Figure 2 Global food profile is significantly associated with the microbiota profile based on the co-inertia analysis (CIA) which investigates the relationship between the two datasets

A) shows the CIA of the FFQ PCA and microbiota PCA where the arrows relate the position of the samples in the FFQ dataset in relation to the microbiota dataset. **B)** shows the FFQ item category associated with the visualised trends. Green dots represent fruits and vegetables, orange represents grains, cereals and bread, brown represents meats, cyan represents fish, yellow represents dairy products, blue represents sweets, cakes and alcohol and grey represents vitamins, minerals and tea. The food items on the most extreme ends are labelled. **C)** shows the microbial taxa at family level associated with visualised trends. The taxa present at the extreme ends are labelled.

FFQ, Food frequency questionnaire; PCA, Principal component analysis

2.4.4 Identification of significantly differentially abundant taxa in patients with osteopenia and osteoporosis

DESeq2 statistical analysis was used to identify genera that were differentially abundant across the groups with adjustment for BMI and gender (**Figure 3A-B, Supplementary table 5**). In summary, we found that *Escherichia/Shigella* and *Veillonella* were more abundant in subjects with osteopenia compared to those with

osteoporosis. *Actinomyces*, *Eggerthella*, *Clostridium Cluster XlVa* and *Lactobacillus* were more abundant in subjects with osteoporosis compared to the normal BMD group. We did not identify any taxa significantly differentially abundant in osteopenia compared to the normal BMD group. The relative abundance of these taxa are shown in **Figure 3C**.

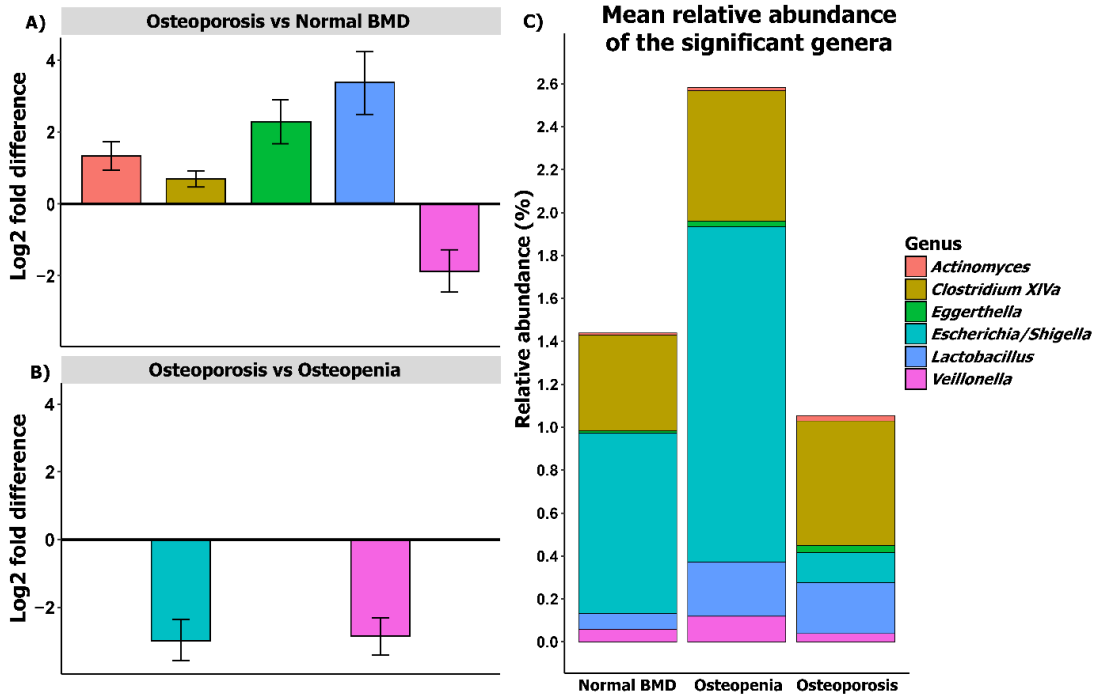


Figure 3 Taxa with differential abundance across the BMD groups
Plot of the Log2 fold difference from the significantly differentially abundant genera in pairwise analysis between the groups from the DESeq2 analysis when the model is adjusted for BMI and gender. Based on the log2 fold difference **A)** shows the genera which are significantly higher in osteoporosis compared to normal BMD. **B)** represents the genera which are significantly more abundant in osteoporosis compared to osteopenia. **C)** represents the relative abundance of the significant genera in the three groups identified in DESeq2.

2.4.5 Alterations at taxonomic levels are not associated with confounding factors

It is well established that many confounding factors may affect the intestinal microbiota (Zhernakova et al., 2016). Therefore, it is important to account for confounders potentially affecting the significant taxa identified. We implemented an in-depth statistical analysis to control for potential cofounders based on a combination of previously published approaches (Jeffery et al., 2016;Zhernakova et al., 2016). Each significant taxon was tested against the confounding meta-data factors as outlined in the Methods section. A total of 29 factors and the inflammatory disease index were analysed and based on the results of the univariate models (**Supplementary Table 6**), the bivariate models explaining the associations with each significant genus were generated (**Supplementary Table 7**).

Significant associations with the different significantly differentially abundant genera were explained by a range of factors including diet, frailty variables, levels of physical activity, medications, weight, BMI, gender, and bone density measurements including the osteopenic and osteoporotic groups **(Supplementary table 6)** based on the univariate models. Based on the bivariate models, five of the six previously identified genera remained significantly differentially abundant after adjustment for known confounding factors **(Supplementary table 7)**. The inflammatory disease index did not show any significant association with these significant taxa in the bivariate models. *Lactobacillus* abundance was not significantly associated with any of the bone density measurements in the univariate and bivariate models unless BMI was included in the model and therefore was no longer considered.

Our analysis shows that BMI is significantly associated with AP spine BMD measures but not with lowest neck of femur BMD values (**Supplementary table 8a,**

Supplementary table 9a). The removal of the effect of BMI, medications and vitamin D levels (**Supplementary table 8b-e, Supplementary table 9b-e**) retained all but two of the results with *Clostridium XlVa* and *Veillonella* losing significance (**Supplementary table 8f, Supplementary table 9f)**.

2.5 Discussion

This is the largest study to-date to investigate associations between the microbiota and reduced bone density in a human cohort including individuals suffering from osteopenia and osteoporosis. We have identified significant associations between different gut microbial genera and reduced bone density in this well-characterised cohort. Extensive investigation by considering the potential influence of various confounders clearly established that the taxonomic differences observed are not explained by the confounders.

It has been observed that the microbiome field suffers from a proliferation of small datasets that show associations of the microbiome with particular diseases or states, without the ability to adequately control for confounding variables. Here we show that global alterations in the gut microbiota are associated with BMD measures, and these interactions explain a similar amount of variance compared to other known microbiota associated diseases and disorders. This confirms our hypothesis of the association of the gut microbiota alterations with a reduction in BMD in the elderly.

Diseases, disorders and medical conditions are associated with smaller effect sizes compared to medications (Falony et al., 2016;Zhernakova et al., 2016). In-depth analysis of confounding variables revealed that bisphosphonate and calcium supplements show no significant association with the global microbiota profile. This is consistent with previous reports that bisphosphonates are not significantly associated with gut microbiota markers and the evidence for microbiota alteration in association with calcium intake is weak (Jackson et al., 2018). We identified six individual gut microbial taxa that may affect bone metabolism. This modest result contrasts with a small cohort study which identified a large number of alterations associated osteoporosis and osteopenia patients in the microbiota at the global and

genus level (Wang et al., 2017) The lack of replication of these global alterations in this cohort shows the importance of adequate sample sizes and controlling for multiple testing when investigating possible new associations.

A loss of microbiota diversity is associated with a wide range of disease states and it is widely considered as an important indicator of health associated microbiota. Within this context, the lack of significant differences in the within-sample diversity measures is interesting. However, it has been observed previously that despite loss of commensal population with the elderly microbiota and noticeable differences in microbiome composition and other host associated factors (e.g. inflammation, dietary patterns), there was no significant observable difference in overall diversity in ageing individuals (Bian et al., 2017) and between frail and non-frail elderly individuals(Jeffery et al., 2016).

The taxa identified resonate well with the bone density-microbiome literature. *Actinomyces* abundance in the osteoporosis group here concords findings that *Actinomyces* is involved in the development of bisphosphonate-related osteonecrosis of the jaw (Arranz Caso et al., 2012) and it has been proposed that prolonged courses of anti-microbial therapy targeting this organism may lead to better clinical outcomes (De Ceulaer et al., 2014). The increase in *Clostridium XlVa* in the osteoporotic group represents a means by which the gut microbiota may influence bone state acting through several differentiating mechanisms (Lopetuso et al., 2013). *Clostridium XlVa* induces accumulation and differentiation of T-reg cells, which in turn are responsible for bone homeostasis (Bozec and Zaiss, 2017). *Clostridium XlVa* is an important producer of butyrate, a short chain fatty acid (SCFA) known to stimulate bone formation (Lucas et al., 2018). Further functional analysis of this group

of microorganisms may provide insight into how the gut microbiota affects BMD through modulation of the host's immune system and metabolism.

Vitamin D receptor polymorphisms are associated with increased osteoporotic fracture risk (Wu et al., 2016). The increase in *Eggerthella* abundance in the osteoporotic group is of interest, as absence of the vitamin D receptor leads to increased *Eggerthella* abundance and other unfavourable alterations in the intestinal microbiota in murine models (Jin et al., 2015). The current investigation also found that vitamin D concentration is associated with a decrease in the relative abundance of *Escherichia/Shigella* (**Supplementary Table 6 and 7**) mirroring other findings looking at vitamin D supplementation (Bashir et al., 2016). The high relative abundance of this genus in osteopenic but not in osteoporotic patients may be partially due to the greater use of oral vitamin D supplementation among the patients with osteoporosis.

A number of microbes belonging to the phylum Firmicutes are known metabolizers of isoflavone diadzin to equol which is an oestrogen analogue (Rafii, 2015). This includes species from the genus *Veillonella*, which we have observed to be decreased in osteoporotic patients. This suggests that a reduction in *Veillonella* would lead to lower production of equol which in turn leads to a lack of inhibition of bone resorption.

An analysis of the meta-data revealed that diet and BMI were large contributors to variance in the dataset, with BMI being the largest single contributor in line with numerous reports linking gut microbiota with obesity (Falony et al., 2016). Our study investigated and confirmed the effect of these variables that can alter the microbiota as reported by previous studies. These included various medications that have a profound effect on the microbiota profiles such as proton pump inhibitors and

the general term of polypharmacy (Ticinesi et al., 2017;Jackson et al., 2018). Thus the current study corroborates previous reports which show that cumulative medication use has the largest effect size on global microbiota profiles (Falony et al., 2016;Zhernakova et al., 2016). However, neither these alterations nor chronic diseases (Falony et al., 2016;Wang et al., 2018) or anthropometric measures explained the observed microbiota alterations.

The relationships between BMI and BMD and the microbiota is complex. Although, lower BMI has been associated with a higher fracture rate (De Laet et al., 2005), a high amount of fat mass may provide no beneficial effect on bone health (Zhao et al., 2007). Within this study, individuals with a higher BMI tended to have higher BMD which is consistent with the literature (Beck et al., 2009). BMI is known to be associated with microbiota alterations. Our analysis has considered both of these BMI associations. Of the taxa related to BMD, *Lactobacillus* and *Veillonella* were significantly related (p-value <0.05) to both obese category and BMD, whilst *Clostridium XlVa* showed trends of associations with the obese category (p-value <0.1). However, the *Lactobacillus* correlation was not significant without adjustment for BMI and so was considered a false positive. Further analysis showed that with removal of variance associated with BMI and medications from the BMD measures results in *Veillonella* and *Clostridium XlVa* losing significance. Other results were unaffected showing that the associations are independent of BMI. Therefore, the association of *Clostridium XlVa* and *Veillonella* with BMD should be interpreted with caution.

This is the first investigation of the intestinal microbiota in a large well-characterised human adult cohort with respect to BMD, with one previous study having a limited sample size (Wang et al., 2017). Nevertheless, the current study has

certain limitations. Due to the recruitment of individuals through consultant referral, the normal BMD cohort are not truly representative of the general population as highlighted by the high fracture rate in this group. However, a history of fractures was not associated with a detectable alteration in the microbiota and controlling for this variable confirmed the BMD results but did not improve the analysis. Due to the incomplete information of the standalone vitamin supplements, we included serum vitamin D levels to use directly measured concentrations to account for vitamin D. The number of variables that can be tested in the identification of confounding factors through statistical analyses is limited by the sample size. However, this analysis was not dependant on the statistical identification of confounding variables, with the majority of the variables being identified from the literature before the commencement of the analysis with all additional variables being supported by the literature. The reported study is also observational and the association with BMD does not imply direct causation. However, the literature supports the notation that these taxa may have functional links to bone health and this microbial contribution to bone health may represent a modifiable environmental factor in the prevention and treatment of osteoporosis. Despite the limitations discussed, changes in gut bacterial composition with respect to bone health suggest further exploration and mechanistic studies are warranted.

In conclusion, we identified taxa-specific differences in the gut microbiota profiles associated with normal bone mineral density, osteopenic and osteoporotic subjects. These genera could be potential biomarkers and therapeutic targets in high risk cohorts. These differences support the concept that specific genera within the gut exert influence on bone metabolism in the host, subsequently affecting bone health in adulthood.

2.6 Acknowledgements

Acknowledgements

We would like to acknowledge the contribution of Jennifer Connolly. We kindly acknowledge the academic support infrastructure of the APC Microbiome Ireland, University College Cork, Cork.

Competing Interests

None to declare

Funding

This work was partly supported by the Irish Centre for Arthritis Research and Education (ICARE). The authors are supported in part by Science Foundation Ireland (SFI) in the form of a research centre grant (SFI/12/RC/2273). IBJ and MD are funded by an SFI Starting Investigator Research Grant (SIRG) (13/SIRG/2128).

Ethics approval

Clinical Research Ethics Committee of the Cork Teaching Hospitals (CREC)

Contributors MD carried out the bioinformatics analysis, compilation and interpretation of the work, drafting and revising of the manuscript submitted. OC, DMK and MGM contributed to the concept and design of the work, acquisition of data and reviewed the manuscript before submission. EMC carried out the compilation of data, DNA extraction and performed the library preparation for NGS. HN and MN carried out the acquisition and compilation of the data. CM carried out the acquisition of data and reviewed the manuscript before submission. PWOT contributed to the concept and design of the work, interpretation of the work, drafting and revising of the manuscript submitted. FS contributed to the concept and design of the work and revising of the manuscript submitted. IBJ contributed to the concept and design of the work, interpretation of the work, drafting and revising of the manuscript submitted.

2.7 References

1. Ali, T., Lam, D., Bronze, M.S., and Humphrey, M.B. (2009). Osteoporosis in inflammatory bowel disease. *Am J Med* 122**,** 599-604.
2. Arranz Caso, J.A., Flores Ballester, E., Ngo Pombe, S., Lopez Pizarro, V., Dominguez-Mompello, J.L., and Restoy Lozano, A. (2012). [Bisphosphonate related osteonecrosis of the jaw and infection with Actinomyces]. *Med Clin (Barc)* 139**,** 676-680.
3. Baker, J.M., Al-Nakkash, L., and Herbst-Kralovetz, M.M. (2017). Estrogen-gut microbiome axis: Physiological and clinical implications. *Maturitas* 103**,** 45-53.
4. Bashir, M., Prietl, B., Tauschmann, M., Mautner, S.I., Kump, P.K., Treiber, G., Wurm, P., Gorkiewicz, G., Hogenauer, C., and Pieber, T.R. (2016). Effects of high doses of vitamin D3 on mucosa-associated gut microbiome vary between regions of the human gastrointestinal tract. *Eur J Nutr* 55**,** 1479-1489.
5. Beck, T.J., Petit, M.A., Wu, G., Leboff, M.S., Cauley, J.A., and Chen, Z. (2009). Does obesity really make the femur stronger? BMD, geometry, and fracture incidence in the women's health initiative-observational study. *J Bone Miner Res* 24**,** 1369-1379.
6. Bian, G., Gloor, G.B., Gong, A., Jia, C., Zhang, W., Hu, J., Zhang, H., Zhang, Y., Zhou, Z., Zhang, J., Burton, J.P., Reid, G., Xiao, Y., Zeng, Q., Yang, K., and Li, J. (2017). The Gut Microbiota of Healthy Aged Chinese Is Similar to That of the Healthy Young. *mSphere* 2.
7. Bozec, A., and Zaiss, M.M. (2017). T Regulatory Cells in Bone Remodelling. *Curr Osteoporos Rep* 15**,** 121-125.
8. Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., Mcdonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., and Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7**,** 335-336.
9. Charlson, M.E., Pompei, P., Ales, K.L., and Mackenzie, C.R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 40**,** 373-383.
10. Claesson, M.J., Jeffery, I.B., Conde, S., Power, S.E., O'connor, E.M., Cusack, S., Harris, H.M., Coakley, M., Lakshminarayanan, B., O'sullivan, O., Fitzgerald, G.F., Deane, J., O'connor, M., Harnedy, N., O'connor, K., O'mahony, D., Van Sinderen, D., Wallace, M., Brennan, L., Stanton, C., Marchesi, J.R., Fitzgerald, A.P., Shanahan, F., Hill, C., Ross, R.P., and O'toole, P.W. (2012). Gut microbiota composition correlates with diet and health in the elderly. *Nature* 488**,** 178-184.
11. Clemente, J.C., Manasson, J., and Scher, J.U. (2018). The role of the gut microbiome in systemic inflammatory disease. *BMJ* 360**,** j5145.
12. Cole, J.R., Wang, Q., Fish, J.A., Chai, B., Mcgarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., and Tiedje, J.M. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 42**,** D633-642.
13. De Ceulaer, J., Tacconelli, E., and Vandecasteele, S.J. (2014). Actinomyces osteomyelitis in bisphosphonate-related osteonecrosis of the jaw (BRONJ): the missing link? *Eur J Clin Microbiol Infect Dis* 33**,** 1873-1880.

14. De Laet, C., Kanis, J.A., Oden, A., Johanson, H., Johnell, O., Delmas, P., Eisman, J.A., Kroger, H., Fujiwara, S., Garnero, P., Mccloskey, E.V., Mellstrom, D., Melton, L.J., 3rd, Meunier, P.J., Pols, H.A., Reeve, J., Silman, A., and Tenenhouse, A. (2005). Body mass index as a predictor of fracture risk: a meta-analysis. *Osteoporos Int* 16**,** 1330-1338.
15. Dinno, A. (2017). dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums. R package version 1.3.4.
16. Donnelly, S., Doyle, D.V., Denton, A., Rolfe, I., Mccloskey, E.V., and Spector, T.D. (1994). Bone mineral density and vertebral compression fracture rates in ankylosing spondylitis. *Ann Rheum Dis* 53**,** 117-121.
17. Drescher, L.S., Thiele, S., and Mensink, G.B. (2007). A new index to measure healthy food diversity better reflects a healthy diet than traditional measures. *J Nutr* 137**,** 647-651.
18. Duvallet, C., Gibbons, S.M., Gurry, T., Irizarry, R.A., and Alm, E.J. (2017). Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat Commun* 8**,** 1784.
19. Eastell, R., O'neill, T.W., Hofbauer, L.C., Langdahl, B., Reid, I.R., Gold, D.T., and Cummings, S.R. (2016). Postmenopausal osteoporosis. *Nat Rev Dis Primers* 2**,** 16069.
20. Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26**,** 2460-2461.
21. Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K., Kurilshikov, A., Bonder, M.J., Valles-Colomer, M., Vandeputte, D., Tito, R.Y., Chaffron, S., Rymenans, L., Verspecht, C., De Sutter, L., Lima-Mendez, G., D'hoe, K., Jonckheere, K., Homola, D., Garcia, R., Tigchelaar, E.F., Eeckhaudt, L., Fu, J., Henckaerts, L., Zhernakova, A., Wijmenga, C., and Raes, J. (2016). Population-level analysis of gut microbiome variation. *Science* 352**,** 560-564.
22. Godin, G., and Shephard, R.J. (1985). A simple method to assess exercise behavior in the community. *Can J Appl Sport Sci* 10**,** 141-146.
23. Guigoz, Y., and Vellas, B. (1999). The Mini Nutritional Assessment (MNA) for grading the nutritional state of elderly patients: presentation of the MNA, history and validation. *Nestle Nutr Workshop Ser Clin Perform Programme* 1**,** 3-11; discussion 11-12.
24. Guss, J.D., Horsfield, M.W., Fontenele, F.F., Sandoval, T.N., Luna, M., Apoorva, F., Lima, S.F., Bicalho, R.C., Singh, A., Ley, R.E., Van Der Meulen, M.C., Goldring, S.R., and Hernandez, C.J. (2017). Alterations to the Gut Microbiome Impair Bone Strength and Tissue Material Properties. *J Bone Miner Res* 32**,** 1343-1353.
25. Hernlund, E., Svedbom, A., Ivergard, M., Compston, J., Cooper, C., Stenmark, J., Mccloskey, E.V., Jonsson, B., and Kanis, J.A. (2013). Osteoporosis in the European Union: medical management, epidemiology and economic burden. A report prepared in collaboration with the International Osteoporosis Foundation (IOF) and the European Federation of Pharmaceutical Industry Associations (EFPIA). *Arch Osteoporos* 8**,** 136.
26. Hind, K., Oldroyd, B., and Truscott, J.G. (2010). In vivo precision of the GE Lunar iDXA densitometer for the measurement of total-body, lumbar spine, and femoral bone mineral density in adults. *J Clin Densitom* 13**,** 413-417.
27. Hsu, E., and Pacifici, R. (2018). From Osteoimmunology to Osteomicrobiology: How the Microbiota and the Immune System Regulate Bone. *Calcif Tissue Int* 102**,** 512-521.

28. Jackson, M.A., Verdi, S., Maxan, M.E., Shin, C.M., Zierer, J., Bowyer, R.C.E., Martin, T., Williams, F.M.K., Menni, C., Bell, J.T., Spector, T.D., and Steves, C.J. (2018). Gut microbiota associations with common diseases and prescription medications in a population-based cohort. *Nat Commun* 9**,** 2655.
29. Jeffery, I.B., Lynch, D.B., and O'toole, P.W. (2016). Composition and temporal stability of the gut microbiota in older persons. *ISME J* 10**,** 170-182.
30. Jin, D., Wu, S., Zhang, Y.G., Lu, R., Xia, Y., Dong, H., and Sun, J. (2015). Lack of Vitamin D Receptor Causes Dysbiosis and Changes the Functions of the Murine Intestinal Microbiome. *Clin Ther* 37**,** 996-1009 e1007.
31. Jones, M.L., Martoni, C.J., and Prakash, S. (2013). Oral supplementation with probiotic L. reuteri NCIMB 30242 increases mean circulating 25-hydroxyvitamin D: a post hoc analysis of a randomized controlled trial. *J Clin Endocrinol Metab* 98**,** 2944-2951.
32. Kanis, J.A., Mccloskey, E.V., Johansson, H., Cooper, C., Rizzoli, R., Reginster, J.Y., Scientific Advisory Board of the European Society For, C., Economic Aspects Of, O., Osteoarthritis, and The Committee of Scientific Advisors of the International Osteoporosis, F. (2013). European guidance for the diagnosis and management of osteoporosis in postmenopausal women. *Osteoporos Int* 24**,** 23-57.
33. Kanis, J.A., Melton, L.J., 3rd, Christiansen, C., Johnston, C.C., and Khaltaev, N. (1994). The diagnosis of osteoporosis. *J Bone Miner Res* 9**,** 1137-1141.
34. Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., and Glockner, F.O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 41**,** e1.
35. Lopetuso, L.R., Scaldaferri, F., Petito, V., and Gasbarrini, A. (2013). Commensal Clostridia: leading players in the maintenance of gut homeostasis. *Gut Pathog* 5**,** 23.
36. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15**,** 550.
37. Lucas, S., Omata, Y., Hofmann, J., Bottcher, M., Iljazovic, A., Sarter, K., Albrecht, O., Schulz, O., Krishnacoumar, B., Kronke, G., Herrmann, M., Mougiakakos, D., Strowig, T., Schett, G., and Zaiss, M.M. (2018). Short-chain fatty acids regulate systemic bone mass and protect from pathological bone loss. *Nat Commun* 9**,** 55.
38. Magoc, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27**,** 2957-2963.
39. Mahoney, F.I., and Barthel, D.W. (1965). Functional Evaluation: The Barthel Index. *Md State Med J* 14**,** 61-65.
40. Molloy, D.W., Alemayehu, E., and Roberts, R. (1991). Reliability of a Standardized Mini-Mental State Examination compared with the traditional Mini-Mental State Examination. *Am J Psychiatry* 148**,** 102-105.
41. Nobel, Y.R., Cox, L.M., Kirigin, F.F., Bokulich, N.A., Yamanishi, S., Teitler, I., Chung, J., Sohn, J., Barber, C.M., Goldfarb, D.S., Raju, K., Abubucker, S., Zhou, Y., Ruiz, V.E., Li, H., Mitreva, M., Alekseyenko, A.V., Weinstock, G.M., Sodergren, E., and Blaser, M.J. (2015). Metabolic and metagenomic outcomes from early-life pulsed antibiotic treatment. *Nat Commun* 6**,** 7486.
42. Peterson, C.T., Sharma, V., Elmen, L., and Peterson, S.N. (2015). Immune homeostasis, dysbiosis and therapeutic modulation of the gut microbiota. *Clin Exp Immunol* 179**,** 363-377.

43. Pytlik, M., Folwarczna, J., and Janiec, W. (2004). Effects of doxycycline on mechanical properties of bones in rats with ovariectomy-induced osteopenia. *Calcif Tissue Int* 75**,** 225-230.
44. Quach, D., Collins, F., Parameswaran, N., Mccabe, L., and Britton, R.A. (2018). Microbiota Reconstitution Does Not Cause Bone Loss in Germ-Free Mice. *mSphere* 3.
45. Rafii, F. (2015). The role of colonic bacteria in the metabolism of the natural isoflavone daidzin to equol. *Metabolites* 5**,** 56-73.
46. Sambrook, P.N., Spector, T.D., Seeman, E., Bellamy, N., Buchanan, R.R., Duffy, D.L., Martin, N.G., Prince, R., Owen, E., Silman, A.J., and Et Al. (1995). Osteoporosis in rheumatoid arthritis. A monozygotic co-twin control study. *Arthritis Rheum* 38**,** 806-809.
47. Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., and Weber, C.F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75**,** 7537-7541.
48. Sjogren, K., Engdahl, C., Henning, P., Lerner, U.H., Tremaroli, V., Lagerquist, M.K., Backhed, F., and Ohlsson, C. (2012). The gut microbiota regulates bone mass in mice. *J Bone Miner Res* 27**,** 1357-1367.
49. Sozen, T., Ozisik, L., and Basaran, N.C. (2017). An overview and management of osteoporosis. *Eur J Rheumatol* 4**,** 46-56.
50. Team, R.C. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing. *Vienna, Austria*.
51. Ticinesi, A., Milani, C., Lauretani, F., Nouvenne, A., Mancabelli, L., Lugli, G.A., Turroni, F., Duranti, S., Mangifesta, M., Viappiani, A., Ferrario, C., Maggio, M., Ventura, M., and Meschi, T. (2017). Gut microbiota composition is associated with polypharmacy in elderly hospitalized patients. *Sci Rep* 7**,** 11102.
52. Wang, J., Wang, Y., Gao, W., Wang, B., Zhao, H., Zeng, Y., Ji, Y., and Hao, D. (2017). Diversity analysis of gut microbiota in osteoporosis and osteopenia patients. *PeerJ* 5**,** e3450.
53. Wang, W., Wang, J., Li, J., Yan, P., Jin, Y., Zhang, R., Yue, W., Guo, Q., and Geng, J. (2018). Cholecystectomy Damages Aging-Associated Intestinal Microbiota Construction. *Front Microbiol* 9**,** 1402.
54. Wu, J., Shang, D.P., Yang, S., Fu, D.P., Ling, H.Y., Hou, S.S., and Lu, J.M. (2016). Association between the vitamin D receptor gene polymorphism and osteoporosis. *Biomed Rep* 5**,** 233-236.
55. Yan, J., Herzog, J.W., Tsang, K., Brennan, C.A., Bower, M.A., Garrett, W.S., Sartor, B.R., Aliprantis, A.O., and Charles, J.F. (2016). Gut microbiota induce IGF-1 and promote bone formation and growth. *Proc Natl Acad Sci U S A* 113**,** E7554-E7563.
56. Yu, Z., and Morrison, M. (2004). Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* 36**,** 808-812.
57. Zhao, L.J., Liu, Y.J., Liu, P.Y., Hamilton, J., Recker, R.R., and Deng, H.W. (2007). Relationship of obesity with osteoporosis. *J Clin Endocrinol Metab* 92**,** 1640-1646.
58. Zhernakova, A., Kurilshikov, A., Bonder, M.J., Tigchelaar, E.F., Schirmer, M., Vatanen, T., Mujagic, Z., Vila, A.V., Falony, G., Vieira-Silva, S., Wang, J.,

Imhann, F., Brandsma, E., Jankipersadsing, S.A., Joossens, M., Cenit, M.C., Deelen, P., Swertz, M.A., Lifelines Cohort, S., Weersma, R.K., Feskens, E.J., Netea, M.G., Gevers, D., Jonkers, D., Franke, L., Aulchenko, Y.S., Huttenhower, C., Raes, J., Hofker, M.H., Xavier, R.J., Wijmenga, C., and Fu, J. (2016). Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* 352**,** 565-569.

2.8 Supplementary Information

2.8.1 Supplementary Methods

Subject recruitments

Patients were screened by telephone and if eligible for study participation, a measurement visit was arranged in the Department of Medicine, Cork University Hospital. All subjects were recruited through referral and stratified post screening. This also allowed for the minimisation of differences in environmental confounding factors between the normal bone density group and the osteoporosis and osteopenia groups compared to if a control group would have been recruited separately through a different methodology.

Bone densitometry assessment

Quality control analysis was performed on the iDXA machine before use on each measurement day. T-score threshold was used to define 3 groups based on their BMD.

Clinical and demographical information

A number of clinical measurements were taken including weight, height, blood pressure, and heart rate using standardized method. Serum blood samples were taken and concentrations of 25-hydroxyvitamin D (25(OH)D) were measured at the Cork Centre for Vitamin D and Nutrition Research using a LC-MS/MS method as described by Cashman *et al.* (2013) (Cashman et al., 2013). Muscle mass was assessed by standard tape measurement of mid-calf and mid-arm circumferences and muscle strength assessed by handheld dynamometer (Jamar, Illinois, USA). A full medical, surgical, and medications history was recorded from each patient including a previous fracture history. Co-morbidities were quantified using the Charlson co-morbidity

index (Charlson et al., 1987). Functional status was assessed using the Barthel index (Mahoney and Barthel, 1965) and mini-mental state examination (MMSE) (Molloy et al., 1991). Patients' physical activity levels were estimated using the Godin-Shepard leisure time physical activity questionnaire (Godin and Shephard, 1985).

Dietary data collection

Dietary habits were assessed by means of a 146-item food frequency questionnaire (FFQ). Participants were asked to record their usual pattern of dietary intake over the previous 3 months. The FFQ was an adapted version of the questionnaire used in the UK arm of the European Prospective Investigation into Cancer (EPIC) study (McKeown et al., 2001), which was based on the original Willet FFQ (Willett et al., 1985). Quality of nutritional intake was assessed using the mini-nutritional assessment (MNA) (Guigoz and Vellas, 1999). Diet quality was represented by the Healthy Food Diversity (HFD) index which was determined from the food frequency of the FFQ dataset based on the publication by Drescher *et al.* (2007) (Drescher et al., 2007).

Molecular methods and bioinformatics

The faecal samples obtained from the participants were stored at -80 °C until processed. Mock communities were not utilized as positive controls in the current study. All samples were processed with the same method and sequenced in a single run. Extraction of genomic DNA from the faecal samples stored in the storage tube was carried out based on modified Yu and Morrison protocol (Yu and Morrison, 2004). The homogenising step was reduced to 60 seconds followed by cooling on ice for 30-60 seconds. The steps were repeated twice more. All other subsequent steps remained same as described in the original protocol. Nuclease free water was used as

a negative control which was processed along with other samples for PCR. As no amplicons were observed in the electrophoresis gel after the amplification step, negative samples were not sent for sequencing. Because faecal samples contain considerably greater microbial biomass than the negative controls, possible reagent contaminant DNA cannot easily outcompete with the amplification template in a faecal DNA preparation (Eisenhofer et al., 2019). It has also been shown that the extraction protocol has a much greater impact on faecal microbiota composition than reagent choices and read counts from negative controls are negligible (Velasquez-Mejia et al., 2018). The V3-V4 region of the 16S ribosomal RNA (rRNA) gene was amplified and sequenced using the following primers:

16S Amplicon PCR Forward Primer (S-D-Bact-0341-b-S-17) = 5'

TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCA

G

16S Amplicon PCR Reverse Primer (S-D-Bact-0785-a-A-21) = 5'

GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATC

TAATCC(Klindworth et al., 2013)

The 2 x 250 paired reads obtained were merged using flash (v1.2.8) (Magoc and Salzberg, 2011) using the following parameters: The average read length is 250. The expected length of merged reads is 460. The standard deviation of length of merged reads is 46. All other parameters were set to default.

The quality filtering of reads and removal of reverse primers were carried out using the QIIME (v 1.9.1) (Caporaso et al., 2010b) pipeline. The split_libraries_fastq.py was set to filter with minimum average quality score set to 20 and truncate_reverse_primer.py with default settings was used to trim. All other settings were set to default. The OTU table was generated as follows using USEARCH

(v8.1) (Edgar, 2010): The unique sequences of length 360-460 were filtered and sorted by length. The sequences were then clustered at 97% (cluster_otus) and chimeras were removed using (uchime_ref) the ChimeraSlayer reference database (microbiomeutil-r20110519)(Haas et al., 2011) to generate representative OTUs. All the reads were mapped against the representative OTUs (usearch_global) and the mapped reads were tabulated in R (v3.4.0)(Team, 2017) to obtain the OTU table. All reported taxa were classified with at least 80% confidence at all levels. Multiple sequence alignment of the representative OTUs were carried out using PyNAST (Caporaso et al., 2010a), and the phylogenetic tree was constructed using FastTree (Price et al., 2009). Calculation of alpha diversity and beta diversity measures were carried out using R and QIIME as follows: using the vegan library (Oksanen et al., 2017)(v2.4.3), a rarefied OTU table was computed by rarefying to the minimum sequencing depth (10613 reads) among the samples. Beta diversity measures (Bray-Curtis) and alpha diversity indexes (chao1, number of unique OTUs (observed richness) and Shannon) were calculated using the rarefied table using the same library. Using the rarefied OTU table and the phylogenetic tree, weighted and unweighted UniFrac distance measure and phylogenetic diversity index were obtained in QIIME.

Analysis of meta-data

Missing values in FFQ dataset were replaced by the median value of that particular food item. Violin plots and stacked bar plots were created using ggplot2(v2.2.1)(Wickham, 2009), RColorBrewer(v1.1.2) (Neuwirth, 2014) and gridExtra(v2.2.1) (Auguie, 2016) to visualise significant meta-data.

Analysis of microbiota data

Bar plots of relative abundance of taxa at phylum level and the top 20 genera based on average abundance across the whole dataset were generated.

To explore the covariance between the FFQ dataset and the microbiota dataset, co-inertia analysis using the ade4 library (v1.7.6) (Dray and Dufour, 2007), was carried out on the Principle Component Analysis (PCA) of log to the base 10 transformed values of the FFQ dataset and microbiota dataset. The first five PCA axes of the microbiota and FFQ dataset were considered for co-inertia analysis. The minimum non-zero value ($1e^{-05}$) was added to the matrix to remove all zeros before log transformation. Significance of the analysis was determined by Monte-Carlo test with 1000 permutations.

PCoA (Principal Coordinates Analysis) was done for visualising the global microbiota profiles in general as explained by any variable and distance dbRDA (distance-based redundancy analysis) was carried out to explore the global profile as explained by the BMD measures. Both PCoA and dbRDA was done using ade4 library on Bray-Curtis distance. Significance of dbRDA was determined by anova.

Body Mass Index (BMI) was categorised into underweight (≤18), normal (>18 and <25), overweight (≥25 and <30) and obese (≥30) for all further analyses.

DESeq2 (v1.16.1) (Love et al., 2014) was used to identify differentially abundant taxa from the microbiota composition table. For the OTU level dataset, DESeq2 analysis was carried out with only group variable. Log2 fold difference of the significant genera based on pairwise comparisons were shown as bar plots created using ggplot2. By default, DESeq2 generates adjusted p-values for only the two groups analysed. As a 3-way comparison was carried out, all of the DESeq2 p-values were adjusted separately using Benjamini-Hochberg method and only those taxa were

selected that passed both DESeq2 adjusted p-values and independently adjusted values.

Identification of meta-data variables associated with beta-diversity

All nominally significant meta-data were analysed to identify their effect size. The effect size was evaluated individually to identify their individual contribution in explaining variance in the global microbiota profiles. Bar-plots and donut plots were made showing the individual effect size, the combined effect size based on pre-defined groups and the overall cumulative effect size. For reproducibility of simulations and permutations, set.seed() function in R was set to 100 for all cases.

Analysis of confounding variables

Analysis of confounding variables was carried out using a univariate general linear mixed model (GLMM) regression with a negative binomial distribution from the library glmmADMB(v0.9.35) (Fournier et al., 2012). Given that R-squared cannot be exactly calculated for mixed models easily. Various methods have been designed recently to approximately obtain a similar value. A pseudo R-squared value is determined for the general linear mixed models in this study using the first formula described by Ben Bolker (https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html#bibliography). All R-squared/effect size values reported are proportional.

Analysis of association of taxa after removal of the effect of BMI, medications and vitamin D levels on BMD measures

A multi-step analysis was carried out to remove the effect of BMI, medications and vitamin D on the BMD measures. At first, we investigated the effect of BMI on the BMD measures by linear regression of BMD to BMI values and determined the associated coefficients and significance using summary. To remove the effect of BMI, medications and vitamin D levels on BMD measures, we generated two multiple regression model one for AP spine BMD and another for Neck of Femur BMD scores. In both these models, the effect of medications (Calcium, Bisphosphonate and Denosumab) and vitamin D levels were investigated. BMI was included in these multiple regression models if it showed significant association with BMD measures in the linear regressions. The residuals were then calculated by removing the effect of the medications investigated and vitamin D levels. Given that the group variable is necessary to show that effect of the medications on BMD measures in the disease groups, the residuals obtained now also had the group effect removed. This effect was added back to the residuals obtained to restore the association with groups and BMD measures excluding BMI, medications and vitamin D levels. To validate this, the values obtained after restoring the effect was modelled against BMI, medications, vitamin D levels and the BMD measures in individual linear regression models and compared against the variations observed between BMD measures to groups. This can be described as follows

Model 1 BMD ~ BMI (If BMI is significant, it is added to the next multiple regression model)

Model 2 BMD ~ BMI + Groups + medication 1 + medication 2 + medication n + vitamin D levels (multiple regression model)

Model 3 BMD ~ Groups (Difference of residuals to BMD from this model is added to residuals of model 2 to restore the effect of groups lost in model 2)

Model 4 Modified residuals ~ Groups (The associated coefficients from model 4 should be similar to model 3)

Model 5 Modified residuals ~ BMD (this result should be near exact fit (coefficient closer to one))

Model 6 Modified residuals ~ BMI/medications/vitamin D levels + Groups (All associations should be with model 4 only and no significance with the other confounders)

Once verified, these residuals without the effect of BMI, medications and vitamin D levels were investigated using univariate GLMMs against the significant taxa identified to determine whether the association of taxa to the disease status were retained or not after removing the effect of BMI, medications and vitamin D levels.

2.8.2 Supplementary Results

Descriptive statistics of the study population

Across all groups, 83% of the participants were women with a mean age of 64.64±5.5 and mean age of male is 63.77±5.76. The BMI of 1% of the participants were underweight while 41% of them were within the normal range, 36% of the participants were overweight and 22% of them were in the obese category. Analysis of the food frequency questionnaire did not reveal any food items to be significantly differentially consumed across the groups. The percentage of missing values in the FFQ dataset was 0.6%. There was no significant difference in diet quality across the groups as defined the HFD metric **(Supplementary table 2)**. Diet is controlled for in all microbiota related results using the HFD metric.

Microbiota characterization

Microbiota profiling and analysis was carried out using 16S rRNA gene amplicon sequencing of V3-V4 region which generated 12,821,961 reads in total from 181 samples. After quality filtering and removal of chimeras, we obtained a table comprising of 4835 OTUs. From 181 participants, a total of 6,512,978 reads were mapped to 4835 OTUs. The average number of reads were 35983±9725.

Identification of significantly differentially abundant taxa in patients with osteopenia and osteoporosis

At the OTU level, significant differences (p-adjusted ≤ 0.05) were observed in the abundance of certain OTUs between the three groups **(Supplementary table 5)**. *Blautia* was more abundant in osteopenia and osteoporosis compared to normal BMD however it was significantly higher in osteopenia compared to osteoporosis also.

Clostridium IV and *Eggerthella* was higher only in the case of osteoporosis compared to normal BMD whereas *Bacteroides* was significantly abundant in normal BMD compared to osteopenia and osteoporosis. *Escherichia/Shigella* was significantly higher in normal BMD and osteopenia compared to osteoporosis but was highest in osteopenia. The results of this analysis were not considered for further analysis due to the sparseness of the dataset.

We applied a DESeq2 model in which we adjusted for fractures; this analysis returned the same results as the above-mentioned model. Two additional taxa, *Akkermansia* and *Klebsiella*, were identified as being more abundant in osteopenia compared to normal BMD subjects but because they were not identified in the earlier model, they were not investigated further.

Alterations at taxonomic levels are not associated with confounding factors

Abundance of *Actinomyces* was positively associated with osteoporosis group and all measures of bone in the univariate model. Bisphosphonate use was not significantly associated when groups category was present, however Coeliac Disease, a prior history of cholecystectomy and total vitamin D levels remained significantly positively associated in the bivariate models. For *Clostridium Cluster XlVa*, an increased abundance was observed with the diseased status. Bisphosphonate and anticoagulant medication and HFD index showed positive significant associated along with the groups variable while Godin score was negatively associated in the bivariate models. Three measures of bone - Groups, AP spine T-score and BMD, were associated with increased abundance of *Eggerthella*. Gender, vitamin D levels, opiod use and previous cholecystectomy were also significantly associated with this genus in the univariate models but only gender and previous cholecystectomy remained

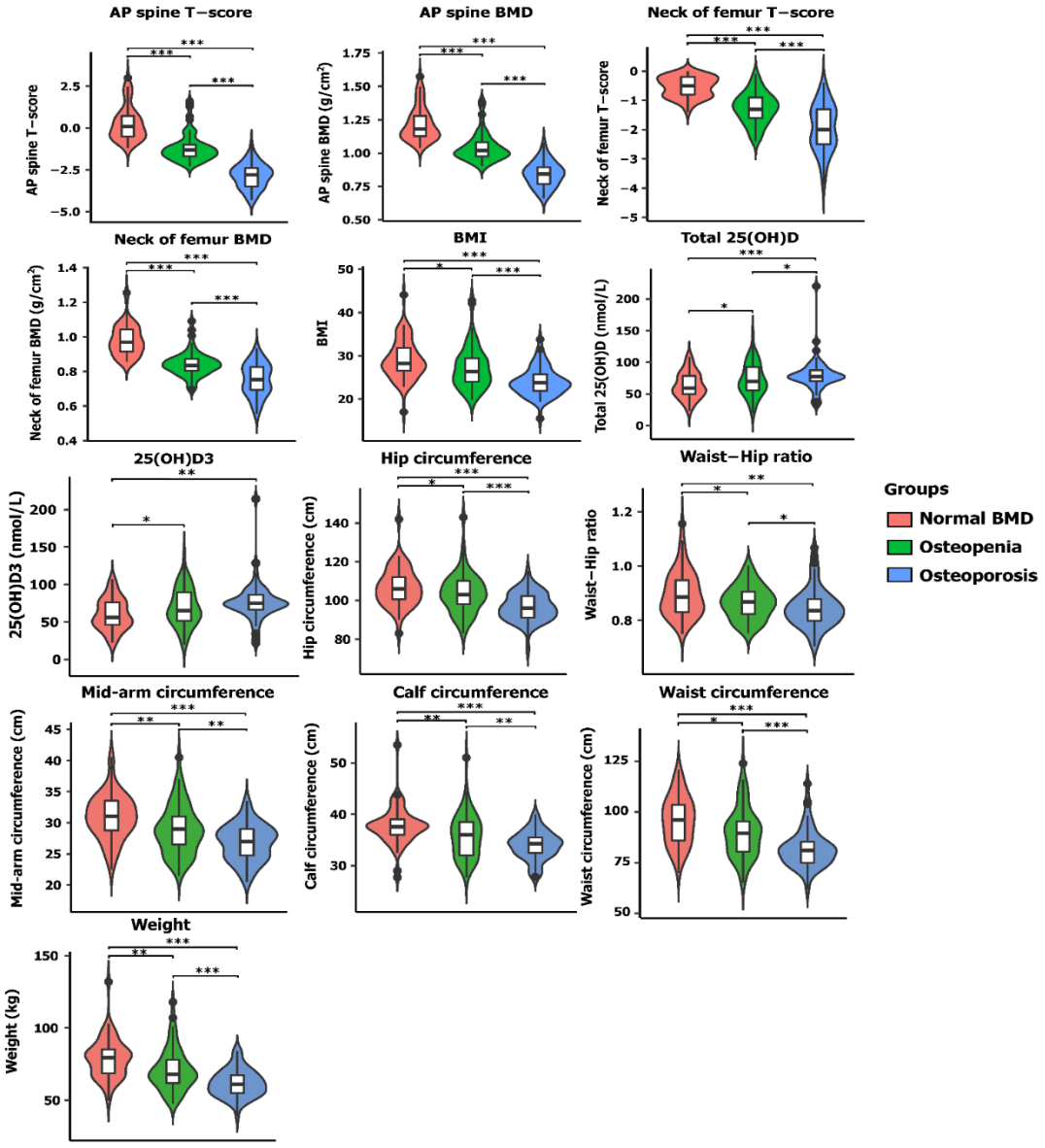
significant in presence of the group category. *Escherichia/Shigella* abundance showed negative associations with vitamin D levels and positively associated with hip circumference measure and appendicectomy. Abundance of this taxon was observed to be lower in osteoporosis and increased in osteopenia. Hip, calf, mid-arm circumference measures and MNA were significantly negatively associated in the bivariate model for *Veillonella* and presence of these confounders showed a non-significant association with disease status. However, after adjustment for calcium and bisphosphonate supplements, BMI, MMSE, medications (anti-epileptics, analgesia (opiods), inhalers and proton pump inhibitors), *Veillonella* was still significantly associated with decreased abundance in osteoporosis and increased in osteopenia. Appendicectomy was also significantly associated after adjustment.

Adjustment for BMI, medications and vitamin D levels

To control for the effect of BMI, medications and vitamin D levels on the bone mineral density, the relationship between these variables and the BMD values were modelled and the residual BMD values generated from the multiple regression models were recorded. These residual BMD values represent the BMD values after adjusting for the effect of BMI, medications (Calcium, Bisphosphonate and Denosumab) and vitamin D levels on the BMD measures (**Supplementary table 8b-9b**). The validation of this removal of effect is described in supplementary table 8c-e and 9c-e. Supplementary table 8c and 9c shows the association of BMD measures to groups categorisation. It can be observed that the Groups categorisation explained 66% of variation with AP spine BMD and 3% variations with neck femur BMD values individuals. The BMD residuals obtained from model 2 and restoring group effect from model 3 shows similar results (**Supplementary table 8d.1, 9d.1**) compared to

supplementary table 8c and 9c. We also observed that the modified BMD residuals explain 97% of variation with BMD measures (**Supplementary table 8d.2, 9d.2**), thus proving that residuals obtained accurately represents the groups and BMD information. To further, validate the removal of confounder effects, we regress this modified residual values against BMI, medications and vitamin D level and show that the variations in the BMD residuals values is not associated at all with BMI, medications or vitamin D levels and any variation explained is attributed the groups status only (**Supplementary table 8e,9e**). After validation of BMD residuals, we used the GLMM regression to show that significant associations are observed with the significant taxa. We observe that the association of the taxa identified to be significant with BMD measures are retained, except for *Clostridium XlVa*, and *Veillonella* where the significance is lost after this adjustment (p-value > 0.05) (**Supplementary table 8f-9f**) and *Lactobacillus* which is not significant in this analysis except for the lowest neck femur residuals.

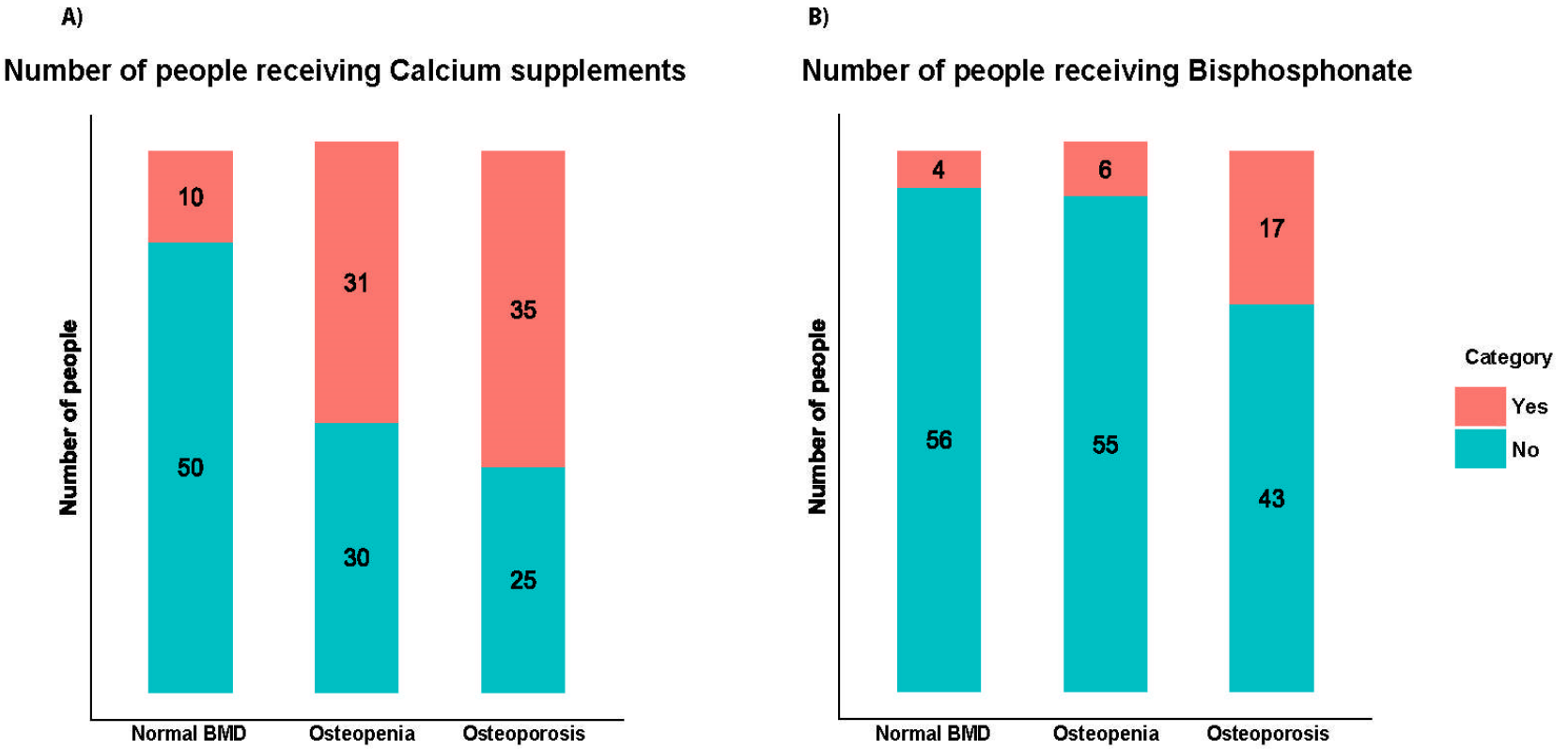
2.8.3 Supplementary Figures



Supplementary figure 1 Variables associated with BMD subject groups

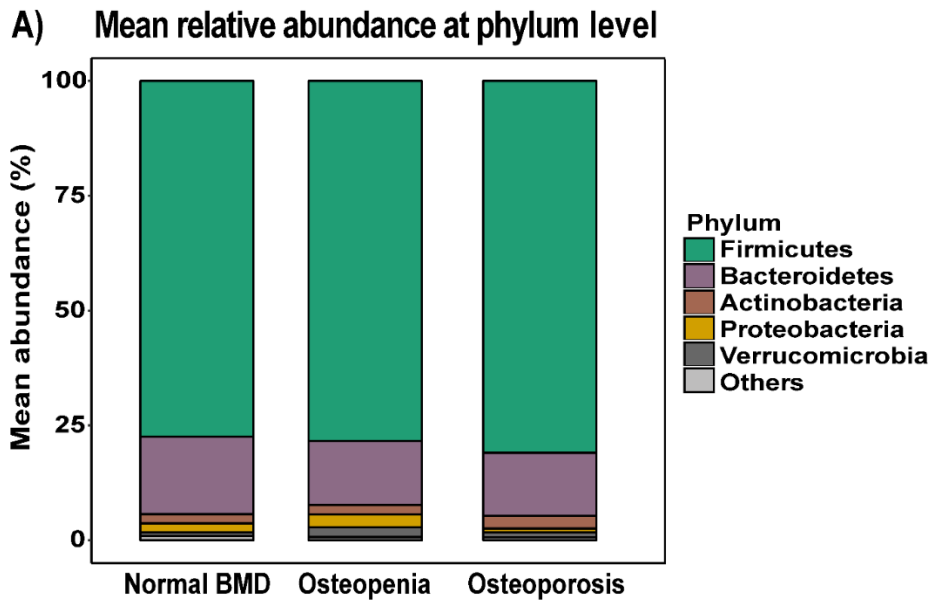
Pairwise comparison of the clinical and BMI data across groups identifies significant alterations in clinical measures in osteoporosis and osteopenia groups compared to the normal BMD group. The violin plots show the distribution of the significant meta-data within each group. Total Vitamin D and Vitamin D3 levels are increased in osteoporosis and osteopenia groups compared to normal BMD, whilst bone mineral density related measurements, BMI and other anthropometric measures are reduced.

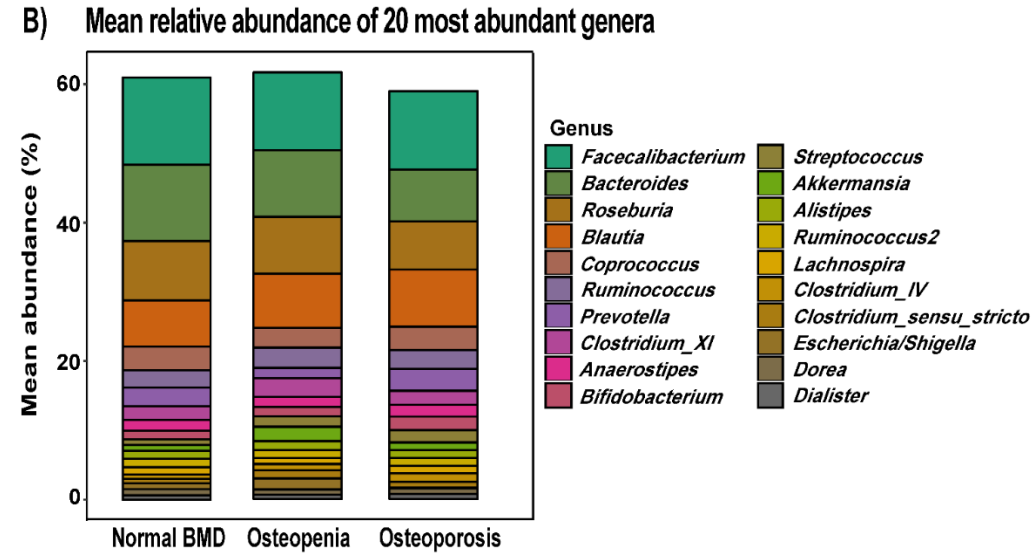
BMI, body mass index; BMD, bone mineral density; AP spine, antero-posterior spine; Total 25(OH)D, total vitamin D; 25(OH)D3, vitamin D3
'***' ≤ 0.0005, '**' ≤ 0.005, '*' ≤ 0.05 (P-adjusted)



Supplementary figure 2 Medications associated with different BMD subject groups

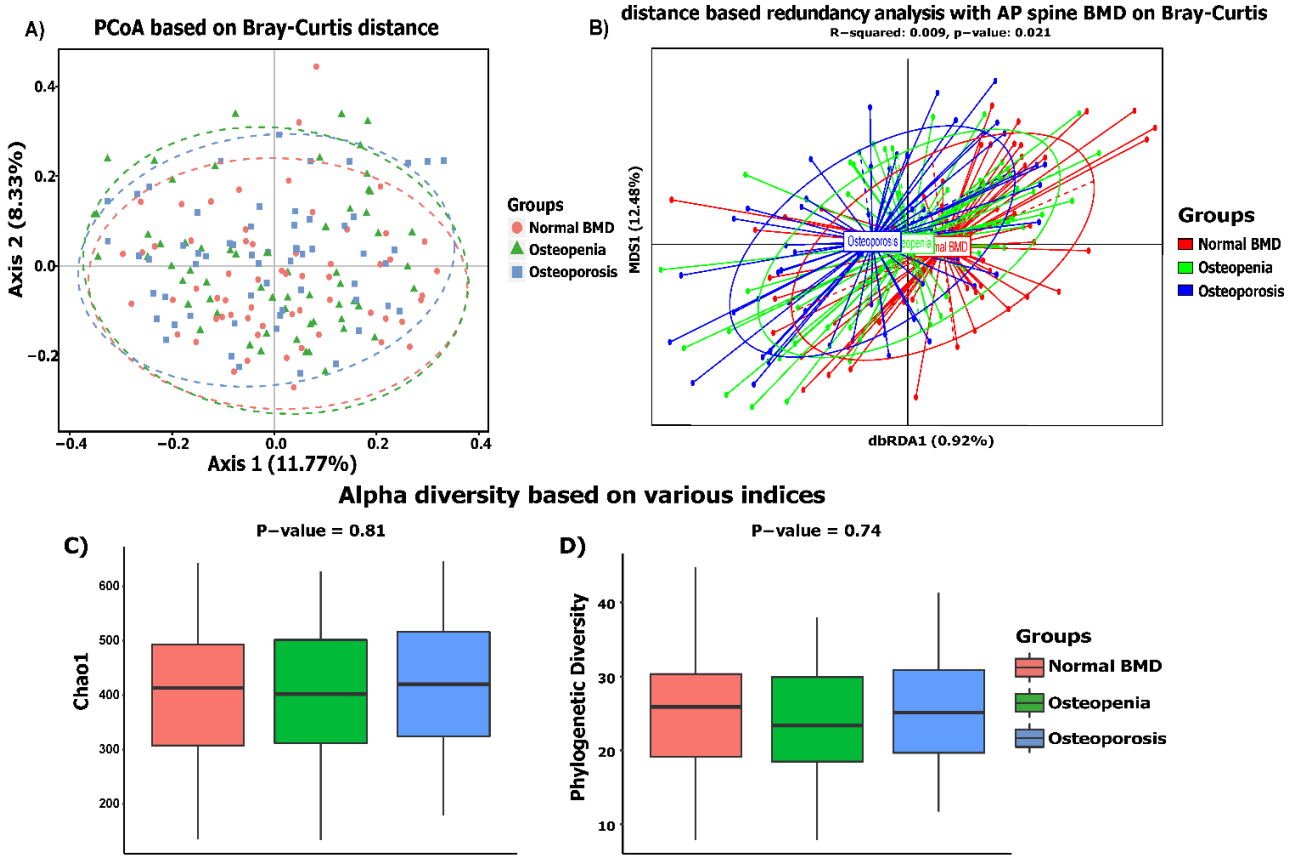
(A) Consumption of Calcium supplements and **(B)** Bisphosphonate medication in the normal BMD controls, osteopenia and osteoporosis individuals.





Supplementary figure 3 Relative abundance of the gut microbiota across different BMD groups

(A) Composition of bacteria at phylum level and **(B)** the 20 most abundant genera in normal BMD, osteopenia and osteoporosis individuals shows similar composition between the groups



Supplementary figure 4 Microbiota α- and β-diversity across the different BMD subject groups

A) Global microbiota profile as observed by unconstrained ordination (PCoA) on Bray-Curtis distance measure. **B)** Constrained ordination i.e. distance-based redundancy analysis (dbRDA) on Bray-Curtis shows global profile best explained by BMD measures. Significant difference in α-diversity between the groups was not observed in **C)** Chao1 and **D)** Phylogenetic distance

2.8.4 Supplementary Tables

Supplementary table 1 Literature citation for all the confounding variables analysed in the dataset

Confounding variables	Literature
Bisphosphonate	Caso et al. 2012. Bisphosphonate related osteonecrosis of the jaw and infection with Actinomyces. Smruti et al. 2014. Oral microbiota and host innate immune response in bisphosphonate-related osteonecrosis of the jaw
Body Mass Index (BMI)	Yun et al. 2017. Comparative analysis of gut microbiota associated with body mass index in a large Korean cohort.
Calcium supplements	Chaplin eta al. 2015. Calcium supplementation modulates gut microbiota in a prebiotic manner in dietary obese mice
Charlson co morbidity	Claesson et. al. 2012. Gut microbiota composition correlates with diet and health in the elderly
Diet (HFD and MNA)	Claesson et. al. 2012. Gut microbiota composition correlates with diet and health in the elderly
Frailty (Barthel Score)	Jackson et. al. 2016. Signatures of early frailty in the gut microbiota. Claesson et. al. 2012. Gut microbiota composition correlates with diet and health in the elderly
Godin Score	Mach and Fuster-Botella 2017. Endurance exercise and gut microbiota: A review
Mini mental score exam (MMSE)	Jeffery et. al. 2015. Composition and temporal stability of the gut microbiota in older persons
vitamin-D	Barbáchano et. al. 2017. The endocrine vitamin D system in the gut. Luthold et al. 2017. Gut microbiota interactions with the immunomodulatory role of vitamin D in normal individuals
Weight	Ley et al. 2006. Microbial ecology: Human gut microbes associated with obesity. Seganfredo et al. 2017. Weight-loss interventions and gut microbiota changes in overwe ight and obese patients: a systematic review

Supplementary table 2 Meta-data analysed in the final dataset. P-value adjustment is based on Benjamini-Hochberg method
Significance: P-adjusted ≤ 0.05

Values mentioned within brackets for the circumference and hand grip measures represents a different sample size from the original

Meta-data	Normal BMD	Osteopenia	Osteoporosis	P-value	P-adjusted	Normal BMD - Osteopenia	Normal BMD - Osteoporosis	Osteopenia - Osteoporosis
Gender (Male/Female)	13/47	7/54	11/49	3.2E-01	5.7E-01			
Age (Years)	63.57±5.73	64.84±5.28	65.07±5.58	2.1E-01	4.5E-01			
Weight (kg)	78.86±13.60	70.96±14.44	61.65±9.44	4.4E-11	6.1E-10	6.4E-04	7.8E-12	1.6E-04
Height (cm)	164.68±8.66	161.25±7.75	160.40±7.56	1.3E-02	5.4E-02			
AP spine T-score	0.28±1.02	-1.16±0.87	-2.86±0.74	1.8E-30	1.3E-28	1.1E-07	2.5E-31	4.4E-11
AP spine BMD (g/cm^2)	1.22±0.13	1.04±0.11	0.84±0.09	8.7E-30	3.0E-28	8.7E-08	1.1E-30	1.4E-10
Neck of femur T-score	-0.54±0.35	-1.27±0.53	-1.95±0.80	1.7E-21	2.9E-20	1.2E-09	5.2E-22	1.1E-04
Neck of femur BMD (g/cm^2)	0.98±0.09	0.84±0.07	0.84±0.68	5.8E-23	1.3E-21	4.7E-11	3.3E-23	2.6E-04
Waist circumference (cm)	95.71±11.95 (13/46)	89.81±12.40 (6/54)	81.81±9.36	2.7E-09	2.7E-08	5.9E-03	6.6E-10	1.4E-04
Hip circumference (cm)	106.71±9.83 (13/46)	103.63±10.45 (6/53)	96.66±7.26	3.7E-08	2.8E-07	4.3E-02	1.8E-08	5.3E-05
Waist-Hip circumference ratio	0.90±0.08 (13/46)	0.87±0.06 (6/53)	0.85±0.07	2.0E-03	9.1E-03	5.0E-02	6.3E-04	4.6E-02
Mid arm circumference (cm)	30.98±3.62 (12/47)	28.85±3.97	26.80±2.91	1.8E-08	1.5E-07	1.1E-03	3.5E-09	2.4E-03
Calf circumference (cm)	37.69±3.73 (11/47)	35.76±4.28	33.93±2.76	7.7E-08	5.3E-07	1.6E-03	1.6E-08	3.5E-03

Hand Grip	27.22±9.03	24.37±8.37 (7/53)	25.88±8.41 (10/48)	1.5E-01	3.7E-01			
25(OH)D2 (nmol/L)	3.19±2.54	2.42±1.63	3.22±3.59	5.4E-01	7.6E-01			
3-epi-25(OH)D3 (nmol/L)	2.97±1.58	3.45±1.97	3.52±2.06	3.0E-01	5.7E-01			
25(OH)D3 (nmol/L)	60.49±20.84	69.98±25.27	75.96±26.43	1.7E-03	8.4E-03	2.7E-02	5.8E-04	7.1E-02
25(OH)D (nmol/L)	63.68±20.57	72.40±25.36	79.18±26.07	1.2E-03	7.0E-03	3.2E-02	3.8E-04	5.0E-02
BMI	29.09±4.57	27.20±4.80	23.96±3.31	6.2E-10	7.1E-09	6.9E-03	1.7E-10	4.5E-05
BMI Categories (number of people)								
Underweight (BMI < 18.5)	1	0	1					
Normal (BMI 18.5 - 24.9)	10	23	41					
Overweight (BMI 25 - 29.9)	27	24	15					
Obese (BMI ≥ 30)	22	14	3					
HFD	0.38±0.10	0.42±0.13	0.41±0.12	1.6E-01	3.8E-01			
MNA (/30)	26.89±2.13	26.86±2.10	26.43±2.47	6.1E-01	8.3E-01			
Systolic BP	137.77±18.05	142.51±18.43	134.37±17.71	5.7E-02	1.9E-01			
Diastolic BP	80.08±9.73	79.87±10.08	76.65±10.30	7.7E-02	2.4E-01			
Pulse	70.58±9.34 (13/46)	71.72±11.36	71.00±10.90	1.0E+00	1.0E+00			
Barthel score	19.68±0.50	19.48±0.87	19.58±0.79	4.2E-01	6.6E-01			
Mini mental score exam	28.48±2.08	28.98±1.57	28.40±2.01	3.1E-01	5.7E-01			
Mini mental score exam ≤ 24 (number of people)	2	1	5					

Godin Leisure time activity score	23.65±22.58	25.56±23.37	24.12±17.71	7.5E-01	
insufficiently active (number of people) ≤ 13	20	20	16		
moderately active (number of people)14-23	15	13	15		9.0E-01
active (number of people) ≥ 24	25	28	29		
Exercises (Number of people)					
none	8	7	6		
Mild	16	16	15		
Moderate	20	18	21		
Mild+Moderate	12	11	13	9.3E-01	1.0E+00
Strenuous	0	1	2		
Moderate+Strenuous	2	3	1		
Mild+Moderate+Strenuous	2	5	2		
Charlson co morbidity (number of people)					
0	37	36	35		
1	13	10	9	6.7E-01	8.7E-01
2	8	13	11		
≥3	2	2	5		
Smoking Status (number of people)					
Never	25	28	25		
Former smoker	25	25	26	9.4E-01	1.0E+00
Current smoker	10	7	9		
Drinking Status (number of people)					
Never	8	5	10		
Former drinker	5	1	3	2.6E-01	5.2E-01
Current drinker	47	55	46		

medications	**Number of medications**	4±2.3	4.4±2.6	4.6±2.6	4.7E-01	6.9E-01				
medications	**Calcium supplements (Yes/No)**	10/50	31/30	35/25	5.2E-06	3.3E-05	2.40E-04	1.80E-05	5.17E-01	
medications	**Bisphosphonate (Yes/No)**	4/56	6/55	17/43	1.5E-03	8.1E-03	7.62E-01	1.20E-02	2.70E-02	
medications	**Analgesia Non-opiod (Yes/No)**	8/52	8/53	8/52	1.0E+00	1.0E+00				
medications	**Analgesia Opiod (Yes/No)**	8/52	7/54	6/54	8.5E-01	9.8E-01				
medications	**Anticoagulants (Yes/No)**	1/59	6/55	1/59	4.1E-02	1.5E-01				
medications	**Antihistamines (Yes/No)**	3/57	2/59	4/56	6.9E-01	8.8E-01				
medications	**Antiplatelets (Yes/No)**	11/49	11/50	9/51	8.7E-01	9.8E-01				
medications	**ARBs ACEi (Yes/No)**	19/41	11/50	11/49	1.2E-01	3.5E-01				
medications	**Cardiovascular Medications (Yes/No)**	13/47	7/54	6/54	1.4E-01	3.6E-01				
medications	**Cholesterol medications (Yes/No)**	28/32	30/31	17/43	4.0E-02	1.5E-01				
medications	**Denosumab (Yes/No)**	1/59	3/58	9/51	1.3E-02	5.4E-02				
medications	**Diuretics (Yes/No)**	8/52	4/57	4/56	3.2E-01	5.8E-01				
medications	**Epilepsy medications (Yes/No)**	6/54	3/58	7/53	3.9E-01	6.4E-01				
medications	**Immunosuppressants (Yes/No)**	3/57	3/58	2/58	8.8E-01	9.8E-01				
medications	**Inhalers (Yes/No)**	7/53	9/52	12/48	4.4E-01	6.6E-01				
medications	**Iron supplementation (Yes/No)**	1/59	3/58	4/56	4.0E-01	6.4E-01				

medications	**Laxatives (Yes/No)**	2/58	1/60	4/56	3.5E-01	6.0E-01
medications	**Minerals (Yes/No)**	2/58	4/57	1/59	3.7E-01	6.1E-01
medications	**Mood Disorders (Yes/No)**	8/52	7/54	10/50	7.0E-01	8.8E-01
medications	**NSAIDs (Yes/No)**	3/57	3/58	6/54	4.4E-01	6.6E-01
medications	**Nutritional Supplements (Yes/No)**	5/55	7/54	7/53	8.0E-01	9.5E-01
medications	**Omega supplementation (Yes/No)**	6/54	2/59	1/59	8.3E-02	2.5E-01
medications	**Other GIT (Yes/No)**	6/54	7/54	7/53	9.5E-01	1.0E+00
medications	**PPIs (Yes/No)**	14/46	14/47	21/39	2.4E-01	5.0E-01
medications	**Sedatives (Yes/No)**	2/58	4/57	4/56	6.6E-01	8.7E-01
medications	**Steroids (Yes/No)**	4/56	6/55	3/57	5.8E-01	8.0E-01
medications	**Thyroid Hormones (Yes/No)**	11/49	14/47	12/48	8.2E-01	9.5E-01
medications	**Vitamins (Yes/No)**	8/52	10/51	7/53	7.5E-01	9.0E-01
medications	**β-blockers (Yes/No)**	9/51	10/51	2/58	4.9E-02	1.7E-01
Chronic diseases	**Appendicectomy (Yes/No)**	13/47	13/48	12/48	9.7E-01	1.0E+00
Chronic diseases	**Bout of depression (Yes/No)**	8/52	3/58	3/57	1.4E-01	1.0E+00
Chronic diseases	**cholecystectomy (Yes/No)**	6/54	5/56	12/48	1.1E-01	3.6E-01
Chronic diseases	**Coeliac (Yes/No)**	9/51	5/56	7/53	5.1E-01	3.2E-01

Chronic diseases	**Diverticulitis (Yes/No)**	5/55	1/60	2/58	1.8E-01	7.3E-01
Chronic diseases	**Inflammatory arthritis (Yes/No)**	3/57	3/58	3/57	1.0E+00	3.9E-01
Chronic diseases	**Reflux (Yes/No)**	3/57	3/58	7/53	2.6E-01	5.2E-01
	Inflammatory Bowel Disease (Yes/No)	2/60	2/59	2/58		
	Multiple Sclerosis (Yes/No)	1/59	1/60	1/59		
	Diabetes Type I (Yes/No)	1/59	0/61	0/60		
	Diabetes Type II (Yes/No)	1/59	4/57	1/59		

BMD, Bone Mineral Density; 25(OH)D2, Vitamin D2; 3-epi-25(OH)D3, epimer of vitamin D3; 25(OH)D3, Vitamin D3; 25(OH)D, Total Vitamin D; BMI, Body Mass Index; HFD, Healthy Food Diversity; MNA, Mini Nutritional Assessment; ARBs/ACEi, angiotensin-receptor blockers/angiotensin converting enzyme inhibitors; NSAIDs, Nonsteroidal anti-inflammatory drugs; GIT, Gastrointestinal; PPIs, Proton Pump Inhibitors

Supplementary table 3 Projected inertia in the first two axis based on PCoA analysis

B-diversity measure	Axis 1	Axis 2
Bray-Curtis	11.77 %	08.33 %
Unweighted UniFrac	14.62 %	04.55 %
Weighted UniFrac	17.45 %	12.67 %

Supplementary table 4 PERMANOVA analysis of various meta-data for its influence on global microbiota profile

	Unweighted		Weighted		Bray-Curtis	
	R2	P-value	R2	P-value	R2	P-value
25(OH)D (nmol/L)	0.005	4.7E-01	0.006	3.4E-01	0.006	3.1E-01
25(OH)D3 (nmol/L)	0.005	4.3E-01	0.007	2.6E-01	0.006	2.4E-01
Age	0.007	1.1E-01	0.007	2.4E-01	0.006	3.5E-01
AP spine BMD (g/cm^2)	0.006	3.8E-01	0.010	**4.0E-02**	0.009	**2.1E-02**
AP spine T.score	0.006	3.5E-01	0.010	**5.0E-02**	0.009	**2.5E-02**
Barthel Score	0.008	6.2E-02	0.010	5.5E-02	0.007	1.5E-01
BMI category	0.018	1.9E-01	0.026	**3.0E-02**	0.020	8.5E-02
Calf circumference (cm)	0.005	6.5E-01	0.004	7.3E-01	0.005	5.8E-01
Carlson Co-morbidity	0.006	3.1E-01	0.010	5.4E-02	0.007	2.2E-01
Gender	0.005	5.1E-01	0.010	**4.2E-02**	0.009	**3.0E-02**
Godin Leisure time activity score	0.007	8.9E-02	0.007	2.2E-01	0.006	2.8E-01
Groups	0.010	7.9E-01	0.015	1.0E-01	0.012	3.5E-01
Hand Grip	0.006	2.5E-01	0.008	1.3E-01	0.008	5.3E-02
Height (cm)	0.007	1.4E-01	0.006	3.4E-01	0.006	2.3E-01

	HFD	0.008	**4.7E-02**	0.008	1.4E-01	0.009	**3.0E-02**
	Hip circumference (cm)	0.006	3.1E-01	0.007	2.7E-01	0.009	5.1E-02
	Lowest Neck of femur BMD (g/cm^2)	0.006	3.8E-01	0.011	**2.8E-02**	0.007	1.5E-01
	Lowest Neck of femur T.score	0.006	1.9E-01	0.007	1.8E-01	0.006	3.0E-01
	Mid-arm circumference (cm)	0.005	5.7E-01	0.005	5.0E-01	0.007	1.5E-01
	MMSE	0.007	9.1E-02	0.011	**3.8E-02**	0.008	**4.4E-02**
	MNA	0.011	**8.0E-03**	0.009	8.6E-02	0.009	**2.5E-02**
	Waist circumference (cm)	0.009	**2.6E-02**	0.011	**3.4E-02**	0.011	**7.0E-03**
	Waist-Hip ratio	0.010	**1.6E-02**	0.012	**1.7E-02**	0.010	**1.5E-02**
	Weight (kg)	0.006	3.7E-01	0.005	4.8E-01	0.007	1.1E-01
Chronic disease	**Appendicectomy (38)**	0.008	**3.9E-02**	0.007	2.1E-01	0.006	2.1E-01
Chronic disease	**cholecystectomy (23)**	0.008	5.1E-02	0.011	**3.9E-02**	0.009	**3.7E-02**
Chronic disease	**Coeliac (21)**	0.008	7.4E-02	0.013	**2.3E-02**	0.011	**6.0E-03**
Chronic disease	**Depression (14)**	0.006	3.4E-01	0.006	3.6E-01	0.007	1.6E-01
Chronic disease	**Diverticulitis (8)**	0.006	2.4E-01	0.006	3.5E-01	0.006	3.5E-01
Chronic disease	**Inflammatory arthritis (9)**	0.008	6.2E-02	0.005	5.0E-01	0.007	1.3E-01
Chronic disease	**Reflux (13)**	0.007	1.7E-01	0.011	**3.0E-02**	0.008	6.2E-02
medications	**Analgesia Non-opiod (24)**	0.006	1.9E-01	0.007	2.3E-01	0.007	1.2E-01
medications	**Analgesia Opiod (21)**	0.009	**1.6E-02**	0.007	2.4E-01	0.008	6.9E-02
medications	**Anticoagulants (8)**	0.013	**2.0E-03**	0.014	**1.4E-02**	0.011	**1.0E-02**
medications	**Antihistamines (9)**	0.006	1.8E-01	0.005	4.9E-01	0.005	6.4E-01
medications	**Antiplatelets (31)**	0.005	7.5E-01	0.006	2.9E-01	0.007	1.9E-01
medications	**ARBs/ACEi (41)**	0.005	5.1E-01	0.010	6.7E-02	0.007	1.0E-01
medications	**β-blockers (21)**	0.005	5.4E-01	0.004	6.9E-01	0.005	5.0E-01
medications	**Bisphosphonate (27)**	0.005	4.1E-01	0.004	6.0E-01	0.005	4.5E-01
medications	**Calcium supplements (76)**	0.005	6.0E-01	0.010	5.4E-02	0.007	1.9E-01
medications	**Cardiovascular Medications (26)**	0.005	7.3E-01	0.002	9.7E-01	0.004	8.8E-01

medications	**Cholesterol medications (75)**	0.005	8.3E-01	0.004	7.1E-01	0.004	8.2E-01
medications	**Denosumab (13)**	0.005	4.2E-01	0.003	8.4E-01	0.004	8.5E-01
medications	**Diuretics (16)**	0.005	4.6E-01	0.005	5.3E-01	0.004	9.1E-01
medications	**Epilepsy medications (16)**	0.007	1.7E-01	0.011	**3.6E-02**	0.008	7.6E-02
medications	**Immunosuppressants (8)**	0.005	4.6E-01	0.009	9.4E-02	0.009	**3.3E-02**
medications	**Inhalers (28)**	0.009	**1.9E-02**	0.004	7.7E-01	0.004	7.9E-01
medications	**Iron supplement (8)**	0.006	2.8E-01	0.007	2.6E-01	0.007	2.1E-01
medications	**Laxatives (7)**	0.007	1.5E-01	0.004	6.7E-01	0.006	2.7E-01
medications	**Minerals (7)**	0.004	9.2E-01	0.003	8.3E-01	0.004	9.3E-01
medications	**Mood Disorders (25)**	0.005	4.9E-01	0.006	3.4E-01	0.007	2.1E-01
medications	**NSAIDs (12)**	0.007	8.5E-02	0.006	3.1E-01	0.005	5.6E-01
medications	**Number of medications**	0.011	**1.0E-02**	0.011	**2.8E-02**	0.010	**1.5E-02**
medications	**Nutritional Supplements (19)**	0.005	7.3E-01	0.004	7.0E-01	0.005	5.9E-01
medications	**Omega supplement (9)**	0.005	6.3E-01	0.002	9.5E-01	0.003	9.9E-01
medications	**Other GIT (20)**	0.006	3.5E-01	0.005	6.0E-01	0.006	2.4E-01
medications	**PPIs (49)**	0.009	**2.1E-02**	0.008	1.6E-01	0.008	5.6E-02
medications	**Sedatives (10)**	0.006	3.5E-01	0.003	9.0E-01	0.003	9.8E-01
medications	**Steroids (13)**	0.005	5.1E-01	0.005	5.1E-01	0.006	3.2E-01
medications	**Thyroid Hormones (37)**	0.006	3.7E-01	0.002	9.6E-01	0.003	9.8E-01
medications	**Vitamins (25)**	0.005	4.5E-01	0.005	4.6E-01	0.004	8.1E-01

Significant nominal p-values are highlighted in red and bold

Supplementary table 5 DESeq2 analysis on the OTU dataset and the genus level dataset. The genus level dataset is adjusted for BMI and/or Gender

DESeq2 OTU level							Relative abundance					
Osteopenia vs Normal BMD							**Normal BMD**			**Osteopenia**		
	log2FoldChange	**lfcSE**	**pvalue**	**padj**	**Family**	**Genus**	**Mean**	**SD**	**SE**	**Mean**	**SD**	**SE**
OTU_1226	2.778	0.581	1.7E-06	1.0E-03	Lachnospiraceae	*Blautia*	0.023	0.072	0.009	0.482	2.729	0.349
OTU_945	2.019	0.465	1.4E-05	4.1E-03	Lachnospiraceae	*unclassified*	0.140	0.782	0.101	0.140	1.382	0.177
							Normal BMD			**Osteoporosis**		
Osteoporosis vs Normal BMD							**Mean**	**SD**	**SE**	**Mean**	**SD**	**SE**
OTU_115	-3.314	0.976	6.9E-04	2.2E-02	Bacteroidaceae	*Bacteroides*	0.009	0.025	0.003	0.144	1.103	0.142
OTU_1866	-1.248	0.322	1.0E-04	1.3E-02	Lachnospiraceae	*unclassified*	0.099	0.148	0.019	0.078	0.095	0.012
OTU_205	1.075	0.295	2.7E-04	1.6E-02	Lachnospiraceae	*unclassified*	0.037	0.038	0.005	0.076	0.107	0.014
OTU_208	1.984	0.550	3.1E-04	1.6E-02	Ruminococcaceae	*Clostridium_IV*	0.018	0.029	0.004	0.046	0.144	0.019
OTU_214	-3.029	0.846	3.4E-04	1.6E-02	Coriobacteriaceae	*unclassified*	0.017	0.042	0.005	0.004	0.016	0.002
OTU_220	2.532	0.607	3.0E-05	5.8E-03	Coriobacteriaceae	*Eggerthella*	0.010	0.015	0.002	0.034	0.077	0.010
OTU_253	1.606	0.435	2.2E-04	1.6E-02	unclassified	*unclassified*	0.009	0.010	0.001	0.020	0.027	0.004
OTU_54	1.344	0.410	1.1E-03	3.1E-02	Ruminococcaceae	*unclassified*	0.148	0.138	0.018	0.251	0.437	0.056
OTU_570	1.834	0.527	5.0E-04	1.8E-02	Lachnospiraceae	*unclassified*	0.046	0.069	0.009	0.135	0.313	0.040
OTU_650	2.709	0.643	2.5E-05	5.8E-03	Lachnospiraceae	*Blautia*	0.003	0.006	0.001	0.021	0.055	0.007
OTU_700	-2.407	0.683	4.2E-04	1.8E-02	Lachnospiraceae	*unclassified*	0.009	0.031	0.004	0.013	0.038	0.005
OTU_732	-3.322	0.959	5.3E-04	1.8E-02	Bacteroidaceae	*Bacteroides*	0.039	0.154	0.020	0.005	0.020	0.003
OTU_8	-2.293	0.623	2.4E-04	1.6E-02	Enterobacteriaceae	*Escherichia/Shigella*	0.840	3.432	0.443	0.138	0.370	0.048
							Osteopenia			**Osteoporosis**		
Osteoporosis vs Osteopenia							**Mean**	**SD**	**SE**	**Mean**	**SD**	**SE**
OTU_1226	-2.626	0.580	6.0E-06	3.5E-03	Lachnospiraceae	*Blautia*	0.482	2.729	0.349	0.084	0.365	0.047
OTU_8	-2.543	0.621	4.2E-05	1.2E-02	Enterobacteriaceae	*Escherichia/Shigella*	1.563	4.785	0.613	0.138	0.370	0.048

DESeq2 Gender BMI Genus level						Normal BMD			Osteopenia		
Osteopenia vs Normal BMD	**log2FoldChange**	**lfcSE**	**pvalue**	**padj**		**Mean**	**SD**	**SE**	**Mean**	**SD**	**SE**
No significant taxa											
						Normal BMD			**Osteoporosis**		
Osteoporosis vs Normal BMD						**Mean**	**SD**	**SE**	**Mean**	**SD**	**SE**
Actinomyces	1.329	0.396	7.9E-04	1.8E-02		0.010	0.011	0.001	0.026	0.041	0.005
Clostridium_XlVa	0.687	0.221	1.9E-03	2.6E-02		0.445	0.392	0.051	0.580	0.525	0.068
Eggerthella	2.282	0.614	2.0E-04	6.9E-03		0.010	0.015	0.002	0.034	0.077	0.010
Lactobacillus	3.355	0.872	1.2E-04	6.9E-03		0.076	0.344	0.044	0.238	0.800	0.103
Veillonella	-1.867	0.592	1.6E-03	2.6E-02		0.057	0.189	0.024	0.039	0.133	0.017
						Osteopenia			**Osteoporosis**		
Osteoporosis vs Osteopenia						**Mean**	**SD**	**SE**	**Mean**	**SD**	**SE**
Escherichia/Shigella	-2.963	0.611	1.2E-06	4.2E-05		1.563	4.786	0.613	0.138	0.370	0.048
Veillonella	-2.849	0.551	2.3E-07	1.6E-05		0.120	0.430	0.024	0.039	0.133	0.017

lfcSE, logfoldchange Standard error; SD, Standard deviation; SE, Standard error

Supplementary table 6 Univariate analysis of the confounding variables against the significant genera
Significant nominal p-values are highlighted in red and bold

Taxa			AIC	BIC	logLik	estimate	p-value	Number of samples
Actinomyces	Groups	Osteopenia	1024.3	1040.3	-507.2	0.352	1.3E-01	181
		Osteoporosis				0.843	**2.3E-04**	
		AP spine BMD (g/cm^2)	1023.5	1036.3	-507.7	-1.653	**3.5E-04**	181
		AP spine T.score	1023.7	1036.5	-507.8	-0.200	**3.9E-04**	181
		Neck of femur T-score	1019.1	1031.9	-505.5	-0.490	**7.0E-05**	181
		Neck of femur BMD (g/cm^2)	1019.8	1032.6	-505.9	-2.699	**2.0E-04**	181
	BMI category	Obese	1038.9	1058.1	-513.5	0.043	8.7E-01	181
		Overweight				0.075	7.3E-01	
		Underweight				-0.799	4.2E-01	
		Calf circumference (cm)	1026.1	1038.8	-509.0	-0.034	1.9E-01	179
		Hip circumference (cm)	1023.4	1036.1	-507.7	-0.003	8.0E-01	178
		Mid arm circumference (cm)	1032.0	1044.8	-512.0	-0.016	5.5E-01	180
		Waist-Hip circumference ratio	1021.8	1034.6	-506.9	1.648	2.1E-01	178
		Waist circumference (cm)	1026.9	1039.7	-509.5	0.005	5.2E-01	179
		Weight (kg)	1035.3	1048.1	-513.6	-0.004	5.4E-01	181
		Gender Male	1035.6	1048.4	-513.8	-0.058	8.2E-01	181
		25(OH)D (nmol/L)	1027.1	1039.9	-509.5	0.011	**3.9E-03**	181
		25(OH)D3 (nmol/L)	1027.9	1040.7	-510.0	0.010	**5.7E-03**	181
		Calcium supplements	1035.4	1048.2	-513.7	0.108	5.8E-01	181
		Bisphosphonate supplements	1027.0	1039.8	-509.5	0.756	**4.8E-03**	181

		Anticoagulants	1035.0	1047.8	-513.5	-0.426	3.9E-01	181
		Number of medications	1033.9	1046.7	-512.9	0.052	1.8E-01	181
		Epilepsy medications	1034.5	1047.3	-513.3	-0.384	2.7E-01	181
		PPIs	1032.3	1045.1	-512.1	0.397	6.9E-02	181
		Analgesia (Opiod)	1034.7	1047.5	-513.4	-0.301	3.2E-01	181
		Coeliac	1030.2	1043.0	-511.1	-0.818	**1.4E-02**	181
		Cholecystectomy	1022.8	1035.6	-507.4	0.922	**7.8E-04**	181
		Reflux	1033.4	1046.2	-512.7	0.510	1.6E-01	181
		Appendicectomy	1035.5	1048.3	-513.8	-0.092	7.0E-01	181
		Inhalers	1034.8	1047.6	-513.4	0.245	3.7E-01	181
		Barthel score	1035.1	1047.8	-513.5	0.106	4.2E-01	181
		Co morbidity	1035.3	1048.1	-513.6	0.066	5.3E-01	181
		Godin	1035.6	1048.4	-513.8	-0.001	8.3E-01	181
		MMSE	1035.6	1048.4	-513.8	-0.008	8.8E-01	181
		MNA	1033.4	1046.2	-512.7	-0.066	1.3E-01	181
		HFD	1034.8	1047.6	-513.4	0.781	3.5E-01	181

Clostridium_XlVa	**Groups**	**Osteopenia**	2213.8	2229.8	-1101.9	0.343	**6.3E-03**	181
		Osteoporosis				0.297	**1.8E-02**	
		AP spine BMD (g/cm^2)	2219.9	2232.7	-1105.9	-0.140	6.2E-01	181
		AP spine T.score	2220.0	2232.8	-1106.0	-0.010	7.8E-01	181
		Neck of femur BMD (g/cm^2)	2216.7	2229.5	-1104.4	0.163	1.0E-01	181

		Neck of femur T-score	2214.9	2227.7	-1103.4	-0.144	**2.3E-02**	181
		Obese				0.264	5.5E-02	
	BMI category	**Overweight**	2218.7	2237.9	-1103.4	-0.034	7.7E-01	181
		Underweight				-0.257	6.1E-01	
		Calf circumference (cm)	2197.4	2210.1	-1094.7	-0.004	7.3E-01	179
		Hip circumference (cm)	2184.5	2197.3	-1088.3	0.007	1.3E-01	178
		Mid arm circumference (cm)	2208.3	2221.0	-1100.1	0.006	6.3E-01	180
		Waist-Hip circumference ratio	2185.8	2198.6	-1088.9	0.769	3.1E-01	178
		Waist circumference (cm)	2195.1	2207.9	-1093.6	0.007	8.3E-02	179
		Weight (kg)	2220.1	2232.9	-1106.0	0.001	8.3E-01	181
		Gender Male	2215.7	2228.5	-1103.9	-0.304	**2.9E-02**	181
		25(OH)D (nmol/L)	2219.9	2232.7	-1106.0	0.001	6.6E-01	181
		25(OH)D3 (nmol/L)	2220.1	2232.9	-1106.1	0.000	8.5E-01	181
		Calcium supplements	2220.1	2232.9	-1106.0	0.028	8.0E-01	181
		Bisphosphonate supplements	2213.5	2226.3	-1102.7	0.368	**1.2E-02**	181
		Anticoagulants	2210.9	2223.7	-1101.4	0.694	**5.4E-03**	181
		Number of medications	2216.6	2229.4	-1104.3	0.039	6.2E-02	181
		Epilepsy medications	2220.0	2232.8	-1106.0	-0.056	7.7E-01	181
		PPIs	2218.1	2230.9	-1105.1	0.165	1.6E-01	181
		Analgesia (Opiod)	2219.9	2232.7	-1105.9	0.084	6.1E-01	181
		Coeliac	2217.3	2230.1	-1104.6	-0.289	8.0E-02	181
		Cholecystectomy	2220.1	2232.9	-1106.1	0.014	9.3E-01	181

		Reflux	2220.1	2232.9	-1106.1	-0.037	8.5E-01	181
		Appendicectomy	2219.4	2232.2	-1105.7	-0.113	3.8E-01	181
		Inhalers	2219.8	2232.6	-1105.9	0.085	5.6E-01	181
		Barthel score	2220.0	2232.8	-1106.0	0.029	6.8E-01	181
		Co morbidity	2219.6	2232.4	-1105.8	-0.041	4.7E-01	181
		Godin	2215.7	2228.5	-1103.9	-0.005	**2.9E-02**	181
		MMSE	2219.9	2232.7	-1105.9	-0.014	6.1E-01	181
		MNA	2216.2	2229.0	-1104.1	-0.047	**4.8E-02**	181
		HFD	2211.4	2224.2	-1101.7	1.228	**3.4E-03**	181

Eggerthella	**Groups**	**Osteopenia**	970.2	986.2	-480.1	0.964	**8.8E-03**	181
		Osteoporosis				1.179	**1.4E-03**	
		AP spine BMD (g/cm^2)	972.1	984.9	-482.1	-2.245	**1.4E-02**	181
		AP spine T.score	973.6	986.4	-482.8	-0.231	**3.5E-02**	181
		Neck of femur BMD (g/cm^2)	977.1	989.9	-484.5	0.205	4.0E-01	181
		Neck of femur T-score	974.2	987.0	-483.1	-0.403	6.1E-02	181
	BMI category	**Obese**				-0.104	8.0E-01	
		Overweight	978.1	997.3	-483.0	0.187	5.9E-01	181
		Underweight				-72.936	1.0E+00	
		Calf circumference (cm)	968.1	980.8	-480.0	-0.039	2.7E-01	179
		Hip circumference (cm)	961.7	974.4	-476.9	-0.001	9.7E-01	178
		Mid arm circumference (cm)	975.1	987.8	-483.5	-0.049	2.5E-01	180

Waist-Hip circumference ratio	960.5	973.2	-476.2	2.652	2.7E-01	178
Waist circumference (cm)	968.9	981.7	-480.5	0.007	6.0E-01	179
Weight (kg)	976.1	988.9	-484.1	-0.016	1.4E-01	181
Gender Male	974.0	986.8	-483.0	-0.912	**2.6E-02**	181
25(OH)D (nmol/L)	973.9	986.7	-482.9	0.014	**4.7E-02**	181
25(OH)D3 (nmol/L)	973.3	986.1	-482.6	0.015	**3.4E-02**	181
Calcium supplements	976.2	989.0	-484.1	0.419	1.8E-01	181
Bisphosphonate supplements	977.9	990.7	-484.9	-0.192	6.6E-01	181
Anticoagulants	977.8	990.6	-484.9	0.343	6.5E-01	181
Number of medications	976.3	989.1	-484.1	0.073	1.9E-01	181
Epilepsy medications	977.5	990.3	-484.7	0.388	4.7E-01	181
PPIs	978.0	990.8	-485.0	-0.075	8.3E-01	181
Analgesia (Opiod)	972.4	985.2	-482.2	0.997	**3.4E-02**	181
Coeliac	977.4	990.2	-484.7	-0.421	3.8E-01	181
Cholecystectomy	968.5	981.3	-480.2	1.220	**6.0E-03**	181
Reflux	977.8	990.6	-484.9	-0.308	6.1E-01	181
Appendicectomy	978.1	990.9	-485.0	0.035	9.3E-01	181
Inhalers	975.4	988.1	-483.7	0.650	1.2E-01	181
Barthel score	977.2	990.0	-484.6	-0.188	3.7E-01	181
Co morbidity	976.4	989.2	-484.2	-0.225	1.9E-01	181
Godin	977.3	990.1	-484.6	-0.008	3.5E-01	181
MMSE	974.7	987.5	-483.3	-0.160	8.1E-02	181

		MNA	976.0	988.8	-484.0	-0.106	1.6E-01	181
		HFD	976.7	989.5	-484.4	1.519	2.5E-01	181

	Groups	**Osteopenia**	1675.1	1691.1	-832.6	1.765	**2.7E-03**	181
		Osteoporosis				-1.090	**4.0E-02**	
		AP spine BMD (g/cm^2)	1698.7	1711.5	-845.3	2.268	**3.8E-02**	181
		AP spine T.score	1698.4	1711.2	-845.2	0.289	**3.5E-02**	181
		Neck of femur BMD (g/cm^2)	1702.4	1715.2	-847.2	-0.216	7.9E-01	181
		Neck of femur T-score	1697.8	1710.6	-844.9	0.737	**1.5E-02**	181
	BMI category	**Obese**	NULL	NULL	NULL	NULL	NULL	181
		Overweight						
		Underweight						
Escherichia/Shigella		**Calf circumference (cm)**	1678.5	1691.3	-835.3	0.076	2.4E-01	179
		Hip circumference (cm)	1674.6	1687.3	-833.3	0.051	**3.2E-02**	178
		Mid arm circumference (cm)	1685.6	1698.4	-838.8	0.085	1.3E-01	180
		Waist-Hip circumference ratio	1676.9	1689.6	-834.5	-5.494	1.7E-01	178
		Waist circumference (cm)	1684.1	1696.8	-838.0	0.022	3.8E-01	179
		Weight (kg)	1702.1	1714.9	-847.1	0.014	5.3E-01	181
		Gender Male	1701.5	1714.3	-846.7	-0.627	2.9E-01	181
		25(OH)D (nmol/L)	1691.8	1704.6	-841.9	-0.025	**2.1E-04**	181
		25(OH)D3 (nmol/L)	1693.0	1705.8	-842.5	-0.024	**5.9E-04**	181
		Calcium supplements	1702.5	1715.3	-847.2	0.031	9.4E-01	181

		Bisphosphonate supplements	1702.0	1714.8	-847.0	-0.469	4.6E-01	181
		Anticoagulants	1702.5	1715.3	-847.2	0.079	9.4E-01	181
		Number of medications	1702.2	1715.0	-847.1	-0.037	5.7E-01	181
		Epilepsy medications	1699.5	1712.3	-845.8	-1.408	**4.1E-02**	181
		PPIs	1701.9	1714.7	-847.0	0.364	4.8E-01	181
		Analgesia (Opiod)	1702.5	1715.3	-847.2	0.044	9.5E-01	181
		Coeliac	1698.5	1711.3	-845.3	-1.454	**2.0E-02**	181
		Cholecystectomy	1701.6	1714.4	-846.8	0.647	3.7E-01	181
		Reflux	1692.6	1705.4	-842.3	-3.497	**2.6E-06**	181
		Appendicectomy	1695.4	1708.2	-843.7	1.474	**2.8E-02**	181
		Inhalers	1698.4	1711.2	-845.2	-1.297	**2.0E-02**	181
		Barthel score	1701.8	1714.6	-846.9	-0.275	4.2E-01	181
		Co morbidity	1702.3	1715.1	-847.2	0.089	7.1E-01	181
		Godin	1698.1	1710.9	-845.0	0.017	5.1E-02	181
		MMSE	1702.4	1715.2	-847.2	0.029	7.5E-01	181
		MNA	1700.1	1712.9	-846.0	-0.151	1.4E-01	181
		HFD	1697.8	1710.6	-844.9	3.714	**3.7E-02**	181
Lactobacillus	**Groups**	**Osteopenia**	1183.3	1199.3	-586.6	0.836	1.9E-01	181
		Osteoporosis				0.319	6.1E-01	
		AP spine BMD (g/cm^2)	1182.3	1195.1	-587.1	-1.319	3.4E-01	181
		AP spine T.score	1182.3	1195.1	-587.2	-0.157	3.5E-01	181

		Neck of femur BMD (g/cm^2)	1182.3	1195.1	-587.2	-0.643	1.8E-01	181
		Neck of femur T-score	1183.2	1196.0	-587.6	0.008	9.7E-01	181
		Obese				1.347	**3.1E-02**	
	BMI category	**Overweight**	1180.0	1199.2	-584.0	0.654	1.8E-01	181
		Underweight				-15.388	9.8E-01	
		Calf circumference (cm)	1164.1	1176.8	-578.0	0.039	6.6E-01	179
		Hip circumference (cm)	1147.5	1160.3	-569.8	0.039	7.1E-02	178
		Mid arm circumference (cm)	1169.2	1182.0	-580.6	-0.040	6.0E-01	180
		Waist-Hip circumference ratio	1149.4	1162.1	-570.7	3.883	2.8E-01	178
		Waist circumference (cm)	1164.8	1177.6	-578.4	0.040	**4.6E-02**	179
		Weight (kg)	1182.9	1195.7	-587.4	-0.011	5.7E-01	181
		Gender Male	1183.0	1195.8	-587.5	-0.300	6.0E-01	181
		25(OH)D (nmol/L)	1182.9	1195.7	-587.4	0.007	5.6E-01	181
		25(OH)D3 (nmol/L)	1182.7	1195.5	-587.4	0.008	4.8E-01	181
		Calcium supplements	1180.8	1193.6	-586.4	0.751	1.2E-01	181
		Bisphosphonate supplements	1172.4	1185.2	-582.2	1.748	**3.4E-03**	181
		Anticoagulants	1183.0	1195.8	-587.5	-0.539	6.2E-01	181
		Number of medications	1169.0	1181.8	-580.5	0.296	**1.2E-04**	181
		Epilepsy medications	1182.6	1195.4	-587.3	-0.608	4.1E-01	181
		PPIs	1183.1	1195.9	-587.6	-0.128	7.9E-01	181
		Analgesia (Opiod)	1182.4	1195.2	-587.2	-0.661	3.2E-01	181
		Coeliac	1182.9	1195.7	-587.5	0.381	5.9E-01	181

		Cholecystectomy	1182.6	1195.4	-587.3	0.479	4.5E-01	181
		Reflux	1182.6	1195.4	-587.3	-0.694	4.1E-01	181
		Appendicectomy	1179.0	1191.8	-585.5	1.116	5.0E-02	181
		Inhalers	1177.5	1190.3	-584.8	1.435	2.5E-02	181
		Barthel score	1168.8	1181.6	-580.4	-0.640	**1.9E-03**	181
		Co morbidity	1175.1	1187.9	-583.6	0.574	**5.2E-03**	181
		Godin	1183.0	1195.8	-587.5	0.004	6.4E-01	181
		MMSE	1182.9	1195.7	-587.5	-0.061	6.1E-01	181
		MNA	1183.2	1196.0	-587.6	-0.009	9.2E-01	181
		HFD	1177.8	1190.6	-584.9	-5.372	**2.0E-02**	181

Veillonella	**Groups**	**Osteopenia**	1249.8	1265.8	-619.9	1.126	**2.0E-03**	181
		Osteoporosis				-0.196	5.9E-01	
		AP spine BMD (g/cm^2)	1262.4	1275.2	-627.2	1.113	3.0E-01	181
		AP spine T.score	1262.8	1275.6	-627.4	0.106	4.2E-01	181
		Neck of femur BMD (g/cm^2)	1251.0	1263.8	-621.5	-3.804	**7.2E-04**	181
		Neck of femur T-score	1244.8	1257.6	-618.4	-0.858	**5.3E-05**	181
	BMI category	**Obese**	1256.3	1275.5	-622.1	-1.447	**3.3E-04**	181
		Overweight				-0.486	1.6E-01	
		Underweight				-1.838	2.1E-01	
		Calf circumference (cm)	1246.3	1259.1	-619.2	-0.096	**3.0E-03**	179
		Hip circumference (cm)	1224.6	1237.3	-608.3	-0.047	**7.4E-03**	178

Mid arm circumference (cm)	1252.1	1264.9	-622.0	-0.107	**2.1E-03**	180
Waist-Hip circumference ratio	1230.6	1243.3	-611.3	1.150	6.8E-01	178
Waist circumference (cm)	1243.2	1255.9	-617.6	-0.026	7.2E-02	179
Weight (kg)	1260.8	1273.5	-626.4	-0.018	8.4E-02	181
Gender Male	1259.6	1272.4	-625.8	0.902	5.0E-02	181
25(OH)D (nmol/L)	1263.4	1276.2	-627.7	0.002	8.1E-01	181
25(OH)D3 (nmol/L)	1263.4	1276.2	-627.7	0.002	7.4E-01	181
Calcium supplements	1257.1	1269.9	-624.6	-0.818	**8.2E-03**	181
Bisphosphonate supplements	1257.5	1270.3	-624.7	-1.231	**4.4E-03**	181
Anticoagulants	1262.7	1275.5	-627.4	0.615	4.2E-01	181
Number of medications	1263.5	1276.3	-627.7	0.009	8.9E-01	181
Epilepsy medications	1254.3	1267.1	-623.1	-2.164	**8.1E-05**	181
PPIs	1245.4	1258.2	-618.7	1.352	**4.8E-05**	181
Analgesia (Opiod)	1259.9	1272.6	-625.9	-1.058	**2.9E-02**	181
Coeliac	1259.5	1272.3	-625.8	0.954	6.5E-02	181
Cholecystectomy	1262.2	1275.0	-627.1	-0.575	2.2E-01	181
Reflux	1240.9	1253.7	-616.5	2.073	**2.4E-04**	181
Appendicectomy	1243.3	1256.1	-617.7	1.520	**1.1E-04**	181
Inhalers	1259.2	1272.0	-625.6	-0.997	**1.9E-02**	181
Barthel score	1259.8	1272.6	-625.9	-0.548	6.6E-02	181
Co morbidity	1262.0	1274.8	-627.0	0.249	2.2E-01	181
Godin	1263.1	1275.9	-627.5	-0.004	4.8E-01	181

		MMSE	1257.3	1270.1	-624.6	0.246	**6.4E-03**	181
		MNA	1249.7	1262.5	-620.8	-0.194	**4.4E-04**	181
		HFD	1261.7	1274.5	-626.9	-1.571	1.8E-01	181

The intercept represents normal BMI, normal BMD, females and No.
AIC, Akaike Information criterion; BIC, Bayesian information criterion

Supplementary table 7: Bivariate analysis of the groups category controlling for significant confounding variables identified from supplementary table 5 against the significant genera

Taxa			AIC	BIC	logLik	estimate	p-value	Number of samples
Actinomyces	**Groups**	**Bisphosphonate supplements**				0.522	5.7E-02	
		Osteopenia	1022.5	1041.7	-505.2	0.336	1.4E-01	181
		Osteoporosis				0.700	**3.1E-03**	
	Groups	**Coeliac**				-0.656	**4.9E-02**	
		Osteopenia	1022.7	1041.9	-505.4	0.293	2.1E-01	181
		Osteoporosis				0.777	**7.5E-04**	
	Groups	**Cholecystectomy**				0.708	**1.2E-02**	
		Osteopenia	1019.6	1038.8	-503.8	0.335	1.4E-01	181
		Osteoporosis				0.639	**6.6E-03**	
	Groups	**25(OH)D (nmol/L)**				0.008	**4.1E-02**	
		Osteopenia	1022.1	1041.2	-505.0	0.262	2.6E-01	181
		Osteoporosis				0.698	**3.0E-03**	
	Groups	**25(OH)D3 (nmol/L)**				0.007	5.5E-02	
		Osteopenia	1022.6	1041.8	-505.3	0.260	2.6E-01	181
		Osteoporosis				0.707	**2.7E-03**	
Clostridium XlVa	**Groups**	**Bisphosphonate supplements**				0.312	**3.7E-02**	
		Osteopenia	2211.3	2230.5	-1099.6	0.310	**1.3E-02**	181
		Osteoporosis				0.222	8.3E-02	
	Groups	**Anticoagulants**				0.648	**1.1E-02**	
		Osteopenia	2208.4	2227.6	-1098.2	0.250	**4.9E-02**	181
		Osteoporosis				0.301	**1.4E-02**	

	Groups	**Gender**	2212.3	2231.5	-1100.2	-0.265	5.4E-02	181
		Osteopenia				0.315	**1.2E-02**	
		Osteoporosis				0.289	**2.0E-02**	
	Groups	**Godin**	2211.2	2230.4	-1099.6	-0.005	**2.6E-02**	181
		Osteopenia				0.345	**5.6E-03**	
		Osteoporosis				0.294	**1.7E-02**	
	Groups	**HFD**	2210.1	2229.3	-1099.1	1.031	**1.8E-02**	181
		Osteopenia				0.248	5.7E-02	
		Osteoporosis				0.264	**3.3E-02**	
	Groups	**MNA**	2213.0	2232.2	-1100.5	-0.039	9.2E-02	181
		Osteopenia				0.322	**1.0E-02**	
		Osteoporosis				0.272	**3.0E-02**	
Eggerthella	**Groups**	**Gender**	966.8	986.0	-477.4	-1.042	**1.0E-02**	181
		Osteopenia				0.934	**1.0E-02**	
		Osteoporosis				1.270	**5.4E-04**	
	Groups	**Analgesia (Opiod)**	967.9	987.1	-478.0	0.874	5.8E-02	181
		Osteopenia				0.800	**3.0E-02**	
		Osteoporosis				1.096	**2.7E-03**	
	Groups	**Cholecystectomy**	965.7	984.9	-476.8	1.102	**1.8E-02**	181
		Osteopenia				0.939	**9.1E-03**	
		Osteoporosis				0.783	**4.0E-02**	
	Groups	**25(OH)D (nmol/L)**	971.2	990.4	-479.6	0.007	3.3E-01	181
		Osteopenia				0.843	**2.8E-02**	
		Osteoporosis				1.029	**1.0E-02**	
	Groups	**25(OH)D3 (nmol/L)**	971.1	990.3	-479.6	0.007	3.1E-01	181
		Osteopenia				0.823	**3.4E-02**	
		Osteoporosis				1.004	**1.3E-02**	
Escherichia/Shigella	**Groups**	**HFD**	1677.0	1696.2	-832.5	0.469	8.0E-01	181
		Osteopenia				1.719	**5.5E-03**	
		Osteoporosis				-1.080	**4.2E-02**	
	Groups	**Epilepsy medications**	1676.5	1695.7	-832.2	0.548	4.5E-01	181
		Osteopenia				1.851	**1.5E-03**	

		Osteoporosis				-1.156	**2.8E-02**	
	Groups	**Coeliac**				-0.724	2.6E-01	
		Osteopenia	1676.0	1695.2	-832.0	1.510	**1.8E-02**	181
		Osteoporosis				-1.257	**2.1E-02**	
	Groups	**Reflux**				-1.506	5.3E-02	
		Osteopenia	1674.2	1693.4	-831.1	1.662	**6.1E-03**	181
		Osteoporosis				-0.971	8.0E-02	
	Groups	**Appendicectomy**				1.049	**3.1E-02**	
		Osteopenia	1672.0	1691.2	-830.0	1.898	**8.1E-04**	181
		Osteoporosis				-0.672	2.1E-01	
	Groups	**Inhalers**				-0.886	1.0E-01	
		Osteopenia	1674.8	1694.0	-831.4	1.576	**6.8E-03**	181
		Osteoporosis				-1.213	**1.8E-02**	
	Groups	**Hip circumference (cm)**				0.049	**3.5E-02**	
		Osteopenia	1648.8	1667.9	-818.4	2.023	**6.0E-04**	181
		Osteoporosis				-0.784	1.4E-01	
	Groups	**25(OH)D (nmol/L)**				-0.022	**2.2E-03**	
		Osteopenia	1669.7	1688.8	-828.8	1.795	**7.6E-04**	181
		Osteoporosis				-0.585	2.9E-01	
	Groups	**25(OH)D3 (nmol/L)**				-0.021	**4.0E-03**	
		Osteopenia	1670.3	1689.5	-829.1	1.781	**1.0E-03**	181
		Osteoporosis				-0.646	2.5E-01	

Lactobacillus	**Groups**	**Barthel score**				-0.666	**1.8E-03**	
		Osteopenia	1172.5	1191.7	-580.2	-0.058	9.2E-01	181
		Osteoporosis				0.205	7.2E-01	
	Groups	**Bisphosphonate supplements**				1.715	**6.3E-03**	
		Osteopenia	1176.4	1195.5	-582.2	0.042	9.4E-01	181
		Osteoporosis				-0.073	9.1E-01	
	BMI category	**Obese**				1.913	**7.8E-03**	
		Overweight				1.561	**1.9E-02**	
		Underweight	1179.2	1204.8	-581.6	-15.371	9.8E-01	181
	Groups	**Osteopenia**				1.226	6.3E-02	
		Osteoporosis				1.753	**2.9E-02**	

		Number of medications				0.302	**9.4E-05**	
	Groups	**Osteopenia**	1171.2	1190.4	-579.6	0.727	1.8E-01	181
		Osteoporosis				0.635	2.8E-01	
		Co morbidity				0.565	**9.3E-03**	
	Groups	**Osteopenia**	1178.3	1197.4	-583.1	0.507	3.9E-01	181
		Osteoporosis				0.521	3.9E-01	
		HFD				-6.173	**6.3E-03**	
	Groups	**Osteopenia**	1178.3	1197.5	-583.2	1.095	5.6E-02	181
		Osteoporosis				0.743	2.5E-01	
		Waist circumference (cm)				0.045	6.0E-02	
	Groups	**Osteopenia**	1167.7	1186.8	-577.9	0.551	3.9E-01	181
		Osteoporosis				0.666	3.2E-01	

		Bisphosphonate supplements				-0.662	1.7E-01	
	Groups	**Osteopenia**	1250.0	1269.2	-619.0	1.101	**2.4E-03**	181
		Osteoporosis				-0.015	9.7E-01	
		Obese				-1.297	**2.6E-03**	
	BMI category	**Overweight**				-0.246	5.4E-01	
		Underweight	1247.2	1272.8	-615.6	-1.069	4.6E-01	181
	Groups	**Osteopenia**				0.844	**4.5E-02**	
		Osteoporosis				-0.474	2.6E-01	
		Calcium supplements				-0.608	1.1E-01	
	Groups	**Osteopenia**	1249.3	1268.5	-618.6	1.171	**1.3E-03**	181
Veillonella		**Osteoporosis**				0.210	6.4E-01	
		Epilepsy medications				-1.907	**4.0E-04**	
	Groups	**Osteopenia**	1243.7	1262.9	-615.8	1.101	**2.1E-03**	181
		Osteoporosis				-0.134	7.1E-01	
		PPIs				1.371	**5.6E-05**	
	Groups	**Osteopenia**	1234.8	1254.0	-611.4	0.601	9.6E-02	181
		Osteoporosis				-0.805	**3.0E-02**	
		Analgesia (Opiods)				-1.074	**2.3E-02**	
	Groups	**Osteopenia**	1247.7	1266.9	-617.9	1.078	**2.9E-03**	181
		Osteoporosis				-0.283	4.4E-01	
		Reflux	1237.5	1256.7	-612.8	1.761	1.9E-03	181

	Groups	**Osteopenia**				0.737	3.9E-02	
		Osteoporosis				-0.164	6.4E-01	
	Groups	**Appendicectomy**				1.512	**1.2E-05**	
		Osteopenia	1229.4	1248.6	-608.7	1.243	**3.1E-04**	181
		Osteoporosis				-0.003	9.9E-01	
	Groups	**Inhalers**				-0.901	**3.0E-02**	
		Osteopenia	1247.9	1267.1	-618.0	1.140	**1.6E-03**	181
		Osteoporosis				-0.117	7.5E-01	
	Groups	**Calf circumference (cm)**				-0.102	**6.5E-03**	
		Osteopenia	1236.3	1255.4	-612.2	0.760	**4.1E-02**	181
		Osteoporosis				-0.615	1.2E-01	
	Groups	**Hip circumference (cm)**				-0.048	**1.4E-02**	
		Osteopenia	1215.9	1235.0	-601.9	0.703	6.9E-02	181
		Osteoporosis				-0.625	1.2E-01	
	Groups	**Mid arm circumference (cm)**				-0.089	**3.9E-02**	
		Osteopenia	1245.5	1264.7	-616.8	0.646	1.2E-01	181
		Osteoporosis				-0.565	1.6E-01	
	Groups	**MMSE**				0.065	5.7E-01	
		Osteopenia	1251.5	1270.6	-619.7	1.049	**7.1E-03**	181
		Osteoporosis				-0.152	6.9E-01	
	Groups	**MNA**				-0.176	**2.1E-03**	
		Osteopenia	1242.2	1261.4	-615.1	0.572	1.4E-01	181
		Osteoporosis				-0.687	7.3E-02	

The intercept represents normal BMI, normal BMD, females and No.

Supplementary table 8: BMI, medications and vitamin D adjustment on AP spine BMD and association with significant genera

a) Identifying association between BMD and BMI (AP spine BMD ~ BMI)

	Estimate	Std. Error	P-values	R2
BMI	0.018	0.003	**1.9E-10**	0.20

b) Investigating the effect of BMI, medications and vitamin D levels on BMD measures

AP spine BMD ~ BMI + Groups + Calcium supplements + Bisphosphonate + Denosumab + vitamin D total + Calcium supplements * Denosumab * vitamin D total		
	Estimate	**Std. Error**
(Intercept)	0.994	0.074
BMI	0.005	0.002
Osteopenia	-0.158	0.021
Osteoporosis	-0.353	0.024
Calcium supplementsYes	0.037	0.062
BisphosphonateYes	0.014	0.024
DenosumabYes	0.099	0.423
Vitamin D Total	0.001	0.001

The intercept represents normal BMD and No

c) Association of AP spine BMD on Groups (AP spine BMD ~ Groups)

	Estimate	Std. Error	P-values	R2
Osteopenia	-0.17	0.02	**2.6E-15**	0.66
Osteoporosis	-0.38	0.02	**< 2E-16**	

The values of estimate, standard error, p-values and R2(variance explained) in linear regression of Groups variable on AP spine BMD

d) Validating the association of residuals from b) on Groups and BMD measure

d.1) Residual obtained from b) and restoring only AP spine variations ~ Groups

	Estimate	Std. Error	P-values	R2
Osteopenia	-0.17	0.02	**2.5E-16**	0.68
Osteoporosis	-0.38	0.02	**< 2E-16**	

The values of estimate, standard error, p-values and R2(variance explained) in linear regression of restored residuals from model 2 on Groups

d.2) Residual obtained from b) and restoring only AP spine BMD ~ actual AP spine BMD

	Estimate	Std. Error	P-values	R2
AP spine BMD	0.97	0.01	**<2E-16**	0.97

The values of estimate, standard error, p-values and R2(variance explained) in linear regression of restored residuals from model 2 on AP spine BMD

e) Residual obtained from b) and restoring only BMD variations ~ medications and BMI

	Estimate	Std. Error	P-values	R2
BMI	1.2E-17	1.9E-03	1.0E+00	
Osteopenia	-1.7E-01	2.0E-02	**7.4E-16**	0.68
Osteoporosis	-3.8E-01	2.2E-02	**2.9E-40**	
Osteopenia	-1.7E-01	2.0E-02	**3.3E-15**	
Osteoporosis	-3.8E-01	2.1E-02	**1.8E-42**	0.68
Calcium supplementsYes	1.7E-17	1.7E-02	1.0E+00	
Osteopenia	-1.7E-01	2.0E-02	**5.5E-16**	
Osteoporosis	-3.8E-01	2.0E-02	**5.1E-44**	0.68
Vitamin D Total	-1.4E-18	3.3E-04	1.0E+00	
Osteopenia	-1.7E-01	1.9E-02	**3.2E-16**	
Osteoporosis	-3.8E-01	2.0E-02	**1.4E-44**	0.68
DenosumabYes	1.1E-16	3.1E-02	1.0E+00	

Osteopenia	-1.7E-01	1.9E-02	**3.1E-16**	
Osteoporosis	-3.8E-01	2.0E-02	**4.2E-44**	0.68
BisphosphonateYes	9.0E-18	2.3E-02	1.0E+00	

The results of linear regression obtained from modelling the restored residuals of model 2 against BMI, medications and vitamin D levels

f) GLMM with residuals obtained from b) and restoring only AP spine BMD variations to taxa

GLMMs	Estimate	Std. Error	P-values	pseudo-R2
Actinomyces	-1.88	0.46	**5.1E-05**	0.09
Clostridium_XlVa	-0.34	0.29	2.4E-01	0.05
Eggerthella	-2.51	0.92	**6.4E-03**	0.03
Escherichia/Shigella	2.96	1.20	**1.3E-02**	0.02
Lactobacillus	-2.22	1.44	1.2E-01	0.08
Veillonella	0.95	1.12	4.0E-01	0.001

Significant nominal p-values are highlighted in red and bold

Supplementary table 9: BMI, medications and vitamin D adjustment on lowest neck-femur BMD and association with significant genera
a) Identifying association between BMD and BMI (Lowest neck femur BMD ~ BMI)

	Estimate	Std. Error	P-values	R2
BMI	-0.001	0.006	8.2E-01	0.0003

b) Investigating the effect of BMI, medications and vitamin D levels on BMD measures

Lowest neck femur BMD ~ Groups + Calcium supplements + Bisphosphonate + Denosumab + vitamin D total + Calcium supplements * Denosumab * vitamin D total		
	Estimate	**Std. Error**
(Intercept)	1.008	0.129
BMI	-0.164	0.078
Osteopenia	-0.185	0.083

Osteoporosis	0.011	0.228
Calcium supplementsYes	0.163	0.090
BisphosphonateYes	0.329	1.563
DenosumabYes	-0.001	0.002
Vitamin D Total	1.008	0.129

The intercept represents normal BMD and No

c) Association of lowest neck-femur BMD on Groups (Lowest neck-femur BMD ~ Groups)

	Estimate	Std. Error	P-values	R2
Osteopenia	-0.137	0.072	**6.0E-02**	0.03
Osteoporosis	-0.136	0.073	**6.3E-02**	

The values of estimate, standard error, p-values and R2(variance explained) in linear regression of Groups variable on lowest neck femur BMD

d) Validating the association of residuals from b) on Groups and BMD measure

d.1) Residual obtained from b) and restoring only Lowest neck-femur variations ~ Groups

	Estimate	Std. Error	P-values	R2
Osteopenia	-0.137	0.071	**5.7E-02**	0.03
Osteoporosis	-0.136	0.072	**5.9E-02**	

The values of estimate, standard error, p-values and R2(variance explained) in linear regression of restored residuals from model 2 on Group

d.2) Residual obtained from b) and restoring only lowest neck-femur BMD ~ actual lowest neck-femur BMD

	Estimate	Std. Error	P-values	R2
Lowest neck femur BMD	0.971	0.013	**<2E-16**	0.97

The values of estimate, standard error, p-values and R2(variance explained) in linear regression of restored residuals from model 2 on lowest neck-femur BMD

e) Residual obtained from b) and restoring only lowest neck-femur BMD variations ~ medications and BMI

	Estimate	Std. Error	P-values	R2
Osteopenia	-0.137	0.075	**6.9E-02**	
Osteoporosis	-0.136	0.077	**7.7E-02**	0.03
Calcium supplementsYes	0.000	0.064	1.0E+00	
Osteopenia	-0.137	0.072	**6.0E-02**	
Osteoporosis	-0.136	0.074	**6.8E-02**	0.03
Vitamin D Total	0.000	0.001	1.0E+00	
Osteopenia	-0.137	0.072	**5.8E-02**	
Osteoporosis	-0.136	0.074	**6.5E-02**	0.03
DenosumabYes	0.000	0.116	1.0E+00	
Osteopenia	-0.137	0.072	**5.7E-02**	
Osteoporosis	-0.136	0.074	**6.8E-02**	0.03
BisphosphonateYes	0.000	0.085	1.0E+00	

The results of linear regression obtained from modelling the restored residuals of model 2 against BMI, medications and vitamin D levels

f) GLMM with residuals obtained from b) and restoring only lowest neck-femur BMD variations to taxa

GLMMs	Estimate	Std. Error	P-values	pseudo-R2
Actinomyces	-2.552	0.638	**6.4E-05**	0.08
Clostridium_XlVa	0.154	0.099	1.2E-01	0.08
Eggerthella	0.246	0.270	3.6E-01	0.04
Escherichia/Shigella	-0.374	0.598	5.3E-01	0.02
Lactobacillus	-0.851	0.360	**1.8E-02**	0.06
Veillonella	-2.030	1.314	1.2E-01	0.001

Significant nominal p-values are highlighted in red and bold, trends of associations are highlighted in green and bold

2.8.5 Supplementary References

1. Auguie, B. (2016). gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.2.1.
2. Caporaso, J.G., Bittinger, K., Bushman, F.D., Desantis, T.Z., Andersen, G.L., and Knight, R. (2010a). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26**,** 266-267.
3. Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., Mcdonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., and Knight, R. (2010b). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7**,** 335-336.
4. Cashman, K.D., Kiely, M., Kinsella, M., Durazo-Arvizu, R.A., Tian, L., Zhang, Y., Lucey, A., Flynn, A., Gibney, M.J., Vesper, H.W., Phinney, K.W., Coates, P.M., Picciano, M.F., and Sempos, C.T. (2013). Evaluation of Vitamin D Standardization Program protocols for standardizing serum 25-hydroxyvitamin D data: a case study of the program's potential for national nutrition and health surveys. *Am J Clin Nutr* 97**,** 1235-1242.
5. Charlson, M.E., Pompei, P., Ales, K.L., and Mackenzie, C.R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 40**,** 373-383.
6. Dray, S., and Dufour, A.B. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software* 22**,** 1-20.
7. Drescher, L.S., Thiele, S., and Mensink, G.B. (2007). A new index to measure healthy food diversity better reflects a healthy diet than traditional measures. *J Nutr* 137**,** 647-651.
8. Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26**,** 2460-2461.
9. Eisenhofer, R., Minich, J.J., Marotz, C., Cooper, A., Knight, R., and Weyrich, L.S. (2019). Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. *Trends Microbiol* 27**,** 105-117.
10. Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., and Sibert, J. (2012). AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software* 27**,** 233-249.
11. Godin, G., and Shephard, R.J. (1985). A simple method to assess exercise behavior in the community. *Can J Appl Sport Sci* 10**,** 141-146.
12. Guigoz, Y., and Vellas, B. (1999). The Mini Nutritional Assessment (MNA) for grading the nutritional state of elderly patients: presentation of the MNA, history and validation. *Nestle Nutr Workshop Ser Clin Perform Programme* 1**,** 3-11; discussion 11-12.
13. Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S.K., Sodergren, E., Methe, B., Desantis, T.Z., Human Microbiome, C., Petrosino, J.F., Knight, R., and Birren, B.W. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* 21**,** 494-504.
14. Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., and Glockner, F.O. (2013). Evaluation of general 16S ribosomal RNA gene PCR

primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 41**,** e1.
15. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15**,** 550.
16. Magoc, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27**,** 2957-2963.
17. Mahoney, F.I., and Barthel, D.W. (1965). Functional Evaluation: The Barthel Index. *Md State Med J* 14**,** 61-65.
18. Mckeown, N.M., Day, N.E., Welch, A.A., Runswick, S.A., Luben, R.N., Mulligan, A.A., Mctaggart, A., and Bingham, S.A. (2001). Use of biological markers to validate self-reported dietary intake in a random sample of the European Prospective Investigation into Cancer United Kingdom Norfolk cohort. *Am J Clin Nutr* 74**,** 188-196.
19. Molloy, D.W., Alemayehu, E., and Roberts, R. (1991). Reliability of a Standardized Mini-Mental State Examination compared with the traditional Mini-Mental State Examination. *Am J Psychiatry* 148**,** 102-105.
20. Neuwirth, E. (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2.
21. Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., Minchin, P.R., O'hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., and Wagner, H. (2017). vegan: Community Ecology Package. R package version 2.4-3.
22. Price, M.N., Dehal, P.S., and Arkin, A.P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 26**,** 1641-1650.
23. Team, R.C. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing. *Vienna, Austria*.
24. Velasquez-Mejia, E.P., De La Cuesta-Zuluaga, J., and Escobar, J.S. (2018). Impact of DNA extraction, sample dilution, and reagent contamination on 16S rRNA gene sequencing of human feces. *Appl Microbiol Biotechnol* 102**,** 403-411.
25. Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York*.
26. Willett, W.C., Sampson, L., Stampfer, M.J., Rosner, B., Bain, C., Witschi, J., Hennekens, C.H., and Speizer, F.E. (1985). Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* 122**,** 51-65.
27. Yu, Z., and Morrison, M. (2004). Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* 36**,** 808-812.

Chapter III

Investigating the Potential of Biologic Treatment to Alter the Gut Microbiota in Rheumatic Inflammatory Disorders

Mrinmoy Das[1,2], David M. Keohane[3], Owen Cronin[3,4], Amy Moloney[1], Siobhan McCarthy[5], John Bourke[6], John Ryan[6], Sinead Harney[6], Mark Phelan[7], Grainne Murphy[6], Catherine Molloy[6], Paul W. O'Toole[1,2], Fergus Shanahan[1,3], Michael G. Molloy[3], Ian B Jeffery[1]

(Manuscript in preparation)

1. APC Microbiome Ireland, University College Cork, Ireland
2. School of Microbiology, University College Cork, Ireland
3. Department of Medicine, University College Cork, Cork University Hospital, Ireland
4. Rheumatic Diseases Unit, Western General Hospital, Edinburgh, United Kingdom
5. Dermatology, South Infirmary Victoria University Hospital, Cork, Ireland
6. Consultant Rheumatologist, Cork University Hospital, Ireland
7. Consultant, South Infirmary Victoria University Hospital, Cork, Ireland

Contents

Chapter 3

Chapter 3 Investigating the Potential of Biologic Treatment to Alter the Gut Microbiota in Rheumatic Inflammatory Disorders

3.1 Abstract

Gut microbiota plays an important role in the maintenance of human health and modulates and is modulated by the immune system. The microbiota has been implicated as a potential risk factor in immune-mediated diseases. The literature suggests that immunosuppressive drugs are associated with microbiota alterations and that these alterations may be a partial restoration of the gut microbiota towards healthy-like structure. No study to date has investigated the effects of biologics on microbiota composition. This pilot study aims to establish if there are changes in gut microbiota associated with biologics treatment and if such changes are observed, if these changes are disease specific or a general property of biologic associated microbiota-recovery.

Metagenomic sequencing was used to compare the gut microbiota profiles of biologics treatment naive subjects undergoing *de-novo* treatment and long-term biologics treated subjects with established diseases (Rheumatoid arthritis (RA), Psoriatic arthritis (PSA), Ankylosing Spondylitis (AS)) compared with healthy controls. This consisted of a multi-timepoint analysis of treatment naive subjects consisting of 132 samples from 24 subjects representing 9 RA, 7 PSA, 4 AS, 4 Inflammatory Bowel diseases (IBD) subjects before and after biologics treatment with sampling pre-treatment and at 3 and 6 months post-treatment. The established group comprises of 10 AS, 18 PSA and 16 RA samples. Healthy samples included 16 subjects sampled across time-points. These subjects were extensively phenotyped with clinical, dietary and anthropometric information investigated.

Our analysis shows that the immunomodulatory medication may partially restore the microbiome to a health-associated state with the β-diversity of the established treatment groups showing no significant differences from the controls. The treatment naive samples showed increased abundances of distinct taxa that were lower in the controls. The microbiota associated with established diseased groups with long-term biologics treatment has an intermediate structure with both control and disease associated taxa being evident. Key taxonomic markers as identified in other studies were also observed and showed significant correlation with clinical variables. Overall, the results indicate that biologics treatment shifts the microbiota signature towards that of healthy controls like. This study provides incentive for the recruitment and analysis of larger cohorts that will allow for the identification of key taxonomic predictors of treatment responsiveness and remission.

Keywords

arthritis, inflammation, biologics, gut microbiota, medication

3.2 Introduction

Immune associated conditions such as different types of arthritis and Inflammatory Bowel Diseases (IBD) are associated with an inflammatory state across body sites (Baumgart and Sandborn, 2007;Ledingham et al., 2017). The aetiology of these conditions are multi-factorial and can affect anyone from a very young age (Baumgart and Carding, 2007;Mezoff and Dykes, 2015;Simone et al., 2018;Smolen et al., 2018;Veale and Fearon, 2018). Arthritis has been reported to affect approximately 0.001-1.4% of populations worldwide (Dean et al., 2014;Liu et al., 2014;Smolen et al., 2018). Epidemiological studies have shown that rheumatoid arthritis (RA) is increasing in prevalence in women whereas ankylosing spondylitis (AS) is more prevalent in men (Landi et al., 2016;Smolen et al., 2018). Occurrence of psoriatic arthritis (PSA) is known to occur at a very young age (Southwood et al., 1989). The detection of these conditions is dependent on development of symptoms and further aggravation of the disease pathogenesis. Amongst these conditions only RA has been identified to be seropositive which allows early patient's evaluation for autoimmune conditions susceptibility (Smolen et al., 2018). Other diseases (AS or PSA) can remain undetected for long time. Factors known to contribute to the onset of these conditions include immunological factors such as HLA-B27, Th17, Th12, Th23 as well as improper protein folding, smoking and stress (Khor et al., 2011;Simone et al., 2018;Smolen et al., 2018;Veale and Fearon, 2018). These conditions are also associated with various co-morbidities and may be associated with other inflammatory conditions such as Inflammatory Bowel disease (IBD), where a subset of patients develop arthritis (Arvikar and Fisher, 2011). The role of gut microbiota in these conditions has not been fully elucidated but animal models have shown the importance of the microbiota, with a lack of disease development in a germ-free (GF) animals.

The subsequent introduction of gut microbiota into these GF animals resulted in the onset of disease symptoms (Peloquin and Nguyen, 2013;Rogier et al., 2017).

The relationship between the gut microbiota in health and disease is well characterised but represents a highly complex relationship (Clemente et al., 2018). One of the prominent roles of the gut microbiota is the proper development, maturation, and maintenance of the immune system (Adlerberth and Wold, 2009). Alterations in the microbiota profiles, community structure or introduction of certain taxa have been related to T-cell mediated modulation of inflammation (Kho and Lal, 2018;Brown et al., 2019). Production of certain bacterial metabolites, degradation of the intestinal mucin layer, or leakage of pathobionts into the host system can also influence the immune system to trigger an inflammatory response (Chow et al., 2011;Kho and Lal, 2018).

Various non-antibiotic medications have also been associated with alterations in the gut microbiota profiles (Maier et al., 2018). Prominent among them are psychotherapeutics, metformin, supplements like vitamin-D, non-steroidal anti-inflammatory drugs (NSAIDs) and proton-pump inhibitors (PPIs) (Zhernakova et al., 2016;Jackson et al., 2018). Anti-Tumour Necrosis Factor (TNF) biologics used for treating immune associated inflammatory conditions have been identified to increase risk of infection (Atzeni et al., 2012).

Medications like Methotrexate, other Disease modifying anti-rheumatic drugs (DMARDs), and biologics that have strong immune-suppressive properties are used extensively to treat arthritis and IBD. However, only a limited number of studies have highlighted potential changes in the gut microbiota dynamics in patients due to immunosuppressive medications.

The aim of the present pilot study was to investigate whether biologics-treated patients with arthritis and IBD show different microbiota profile dynamics over the course of treatment and recovery. In addition, we aim to uncover the directionality of any such change in the gut microbiota profiles, and key marker features associated with different disease and recovery stages. We hypothesise that the disease-associated gut microbiota profiles demonstrate dysbiosis in the treatment-naive stage which is restored towards the healthy like microbiota with treatment.

3.3 Methods

3.3.1 Subject recruitment and clinical information

Samples were recruited from Cork University Hospital and Irish Centre for Arthritis Research and Education (iCARE). Patients diagnosed with rheumatoid arthritis (RA), psoriatic arthritis (PSA), ankylosing spondylitis (AS) and Inflammatory Bowel Diseases (IBD) were considered for the study. Two different groups of participants were considered. The patients from the first groups were diagnosed with the disease but were not treated with biologics. Samples from this group of patients were collected over three time-points of three-month intervals where the first sample is collected before biologics treatment and the subsequent samples were collected during biologics treatment. The second group of patients recruited were an established disease group that had been on biologics for over a year. Control participants were obtained from the ControlMet cohort of APC Microbiome Ireland, Cork City and were similarly collected over three time-points. All participants provided written consent. Exclusion criteria included known history of alcohol abuse, participation in an investigational drug trial in the 30 days before enrolment, use of antibiotics in the 3 months prior recruitment.

Clinical information included age, sex, anthropometric measures, smoking status, alcohol status, medication, surgery, and other medical history. Effectiveness of therapy was also recorded. It is defined by Bath Ankylosing Spondylitis Disease Activity Index (BADSAI) for AS (Garrett et al., 1994), Disease Activity in Psoriatic Arthritis (DAPSA) for PSA (Schoels et al., 2010), Disease Activity Score of 28 joints (DAS28) for RA (Prevoo et al., 1995), Harvey-Bradshaw index for Crohn's disease (CD) (Harvey and Bradshaw, 1980) and Powell-Tuck index for ulcerative colitis (UC) (Powell-Tuck et al., 1978). Morisky score was used to determine adherence to

medications (Morisky et al., 1986). Population of various immune-cells (WBCs, Monocytes, Neutrophils, Lymphocytes, Basophils and Eosinophils) and levels of C-reactive protein (CRP) were obtained from the blood of the diseased samples. Venous blood samples were collected from the antecubital fossa, where possible. One Greiner Bio-One VACUETTE® K3EDTA and one VACUETTE® Z Serum Sep Clot Activator tube was filled from each participant and centrifuged as per manufacturer's specifications to produce a plasma and serum sample. All haematological (including total and differential white cell count) and biochemical (high sensitivity CRP) analysis was performed in an accredited University Hospital laboratory compliant with international best standards of quality and competence (ISO 15189:2012). The control samples lacked serum information. International Physical Activity Questionnaires (IPAQ) was used to calculate the health-related physical activity of the individuals (Hagstromer et al., 2006). Dietary information was recorded using food frequency questionnaires adapted from the European Prospective Investigation into Cancer (EPIC) study (UK arm) (McKeown et al., 2001). Any missing dietary information for samples was inferred from the same subjects at subsequent time-points by considering the recorded values from these time-points. Any other missing values in the samples were replaced by the median value of that particular food item observed for that group of samples. Healthy Food Diversity (HFD) index was calculated to determine diet quality (Drescher et al., 2007).

3.3.2 Sample collection and DNA extraction

Faecal and blood samples were collected from the subjects. Faecal samples were stored in -80 until further processing. DNA extraction was carried out using modified Yu and Morrison protocol (Yu and Morrison, 2004). Homogenisation was carried out

for 60 seconds followed by 30-60 seconds of cooling on ice. These modified steps were repeated twice more. All other steps remained unchanged.

3.3.3 Sample processing

Preparation of faecal samples for metagenomics whole genome shotgun (mWGS) sequencing was carried out using standard Illumina protocol for paired 150 X 2 MiSeq run. Paired reads obtained were trimmed and quality filtered using Trimmomatic (v.0.38) (Bolger et al., 2014) with the following parameters. The leading and trailing end were trimmed up to 10 bases. Sliding window of 4 bases and average quality of 20 was set. Reads were retained if the trimmed length were at least 75 bases long. Quality filtered reads were then processed with bowtie2 (v.2.2.3) (Langmead and Salzberg, 2012) and mapped against the human genome (GRCh 38) to remove host contamination.

Taxonomic profile was obtained using MetaPhlAn2 (Truong et al., 2015) at default setting. Using HUMAnN2 (v.0.7.1) (Franzosa et al., 2018) with UniRef90 protein database and larger structured MetaCyc pathways mapping file, functional profile was obtained. The datasets were normalised to relative abundance. Eukaryotes, viruses, and archaea were removed from the taxonomic dataset, whereas all UMAPPED, UniRef_unknown and UNINTEGRATED were removed from functional datasets after normalisation to relative abundances.

From the taxonomic dataset, α-diversity indices (Observed species and Shannon) were calculated using the vegan library (v.2.5.4) (Oksanen et al., 2017) in R (v.3.6.0) (Team, 2019). B-diversity defined by Spearman dissimilarity measure was calculated from the taxonomic profiles. Functional diversity index (observed gene

count) was obtained from UniRef gene dataset. Spearman dissimilarity measure was also calculated from the relative abundance of pathways profiles.

3.3.4 Statistical analysis

All statistical analyses described below were carried out in R. Statistical significance is determined by P-adjusted ≤ 0.05 (Benjamini-Hochberg method) or nominal P-value ≤ 0.05 cut-off unless stated otherwise. Species level taxonomic data and non-stratified pathway abundance data were used for all analysis.

3.3.5 Analysis of meta-data

All meta-data values were compared between time-points using paired Wilcox test for any significant changes throughout the duration. To compare disease groups to control, disease samples from timeseries and established groups were compared with Control T0 using Mann-Whitney test. Diseased samples from timeseries groups were compared against the respective disease established groups using Mann-Whitney test to identify alterations associated with disease and treatment progression.

3.3.6 Analysis of global microbiota composition

Alterations in the global composition defined by various dissimilarity measures were analysed using bca (between-class analysis) from ade4 library (v.1.7.13) (Dray and Dufour, 2007) on datasets containing only a particular disease type along with the control samples. Significance and effect size were determined using Monte-Carlo algorithm for bca.

Differences between timepoints were investigated using the beta diversity dissimilarity values as modelled by a linear regression model with adjustment for

observed species. This was done because reduced α-diversity is associated with increased variability in β-diversity (Jeffery et al., 2016). The model is as follows:

Distance between timepoints ~ observed species + Groups

Absolute dissimilarity measures from T2 to T0 and T3 to T0 was calculated for the different disease groups.

An independent dataset by Zhang *et al.* (Zhang et al., 2015) comprising of RA samples was also included in the β-diversity analysis. The Zhang *et al.* dataset was filtered to remove control relatives, as they could not be related back to the clinical data. Amongst the samples in the Posttreatment group's stratification, those groups that were represented by less than 3 biological samples were also excluded from the analysis.

Using paired Wilcox test on α-diversity indices and observed gene count identified significant differences between time-points. These indices from different disease groups were compared against controls (T0) using Mann-Whitney test. The indices from different diseases were also compared between time-points and established groups.

3.3.7 Identification of discriminating taxonomic features

Investigation of the taxonomic profiles from the individual disease groups was carried by comparing against the control T0 samples. The taxonomic datasets from each disease groups were filtered to remove species present below 0.001% abundance and detected in less than 10% of the samples. Identification of differentially abundant taxa was carried out using DESeq2 library (v.1.24.0) (Love et al., 2014). To generate count data. relative abundance data was multiplied by the number of reads mapped to clade-specific genes. This is adapted from the tutorial provided in curatedMetagenomicData

(Pasolli et al., 2017) for analysing MetaPhlAn2 profiles. Each disease group was investigated individually, and significant species were identified based on p-adjusted ≤ 0.05 in each pairwise analysis. Significant taxonomic features identified more than once across all pairwise comparisons were reported. This was done to remove spuriously detected taxa that were not reproduced in any pairwise comparisons. These were visualised using the heatmap.2 functionality from gplots (v.3.0.1.1) using the transformed values derived from the DESeq2 analysis. Heatmap dendrogram was generated using Kendall distance with ward clustering on the transformed values of the reported taxa.

Differential abundance of pathways was carried out using Kruskal-Wallis test between diseased samples and control T0 samples adjusted for multiple testing across pathways. Significant p-adjusted (≤ 0.05) pathways were further investigated using dunn's test (Dinno, 2017).

3.3.8 Association between meta-data and significant taxa

Key meta-data were correlated against significant taxa using Kendall correlation to identify significant association between meta-data and significant taxonomic markers and significance of Kendall's correlation were defined by nominal p-value (≤ 0.05). Correlations between meta-data and significant taxa was visualised using heatmaps. Dendrogram for meta-data and significant taxa was obtained using Euclidean distance on the correlation values using ward.D2 clustering.

3.4 Results

3.4.1 Description of the study population

The present study comprised of 132 samples stratified into their respective groups as detailed in **Supplementary tables 1-2**. The samples comprised of established disease patients and a time-point of biologic treatment that consisted of pre-treatment (biologic naive), and 3 and 6 post-treatment samples. All biologics naive samples were deemed to have active disease by a physician despite being on methotrexate (MTX) and/or non-steroidal anti-inflammatory drugs (NSAIDs). The established samples are represented by their abbreviated disease names whereas the timeseries samples are labelled as abbreviated disease with their respective time-point. IBD samples comprised of 3 CD and 1 UC subjects.

Subjects in both the established and timeseries disease groups were older compared to controls. They had a reduced dietary index (HFD), a smaller stature and a higher BMI than the control group (**Supplementary table 2, Figure 1**). A total of 53% of patients were currently or previously ex-smokers whereas only 31% of the control subjects had a history of smoking. Among the disease subjects 57% of them reported a history of immune disorder in their family.

Statistical analysis of the time-series meta-data showed that the majority of patient characteristics were stable over the timeseries **(Supplementary table 3)** with noticeable exceptions being an increase in HFD and the Morisky score for the RA timeseries and CRP levels between T2 compared to T1 mentioned above.

Levels of CRP were significantly higher between pre-treatment RA (T0) and established RA **(Supplementary table 4, Figure 1B)** but dropped at the T2 timepoint as expected but was not significant at the T3 timepoint.

Immune cells population in blood was observed to be higher in biologic naive RA (T0) than biologic treated established RA with ssignificant differences observed between T0 vs T2 for white blood cells, Neutrophils, Lymphocytes and Monocytes. No immune cell differences reached significance in the other disease states. This may be due to lack of statistical power due to insufficient sample size **(Supplementary table 1)**.

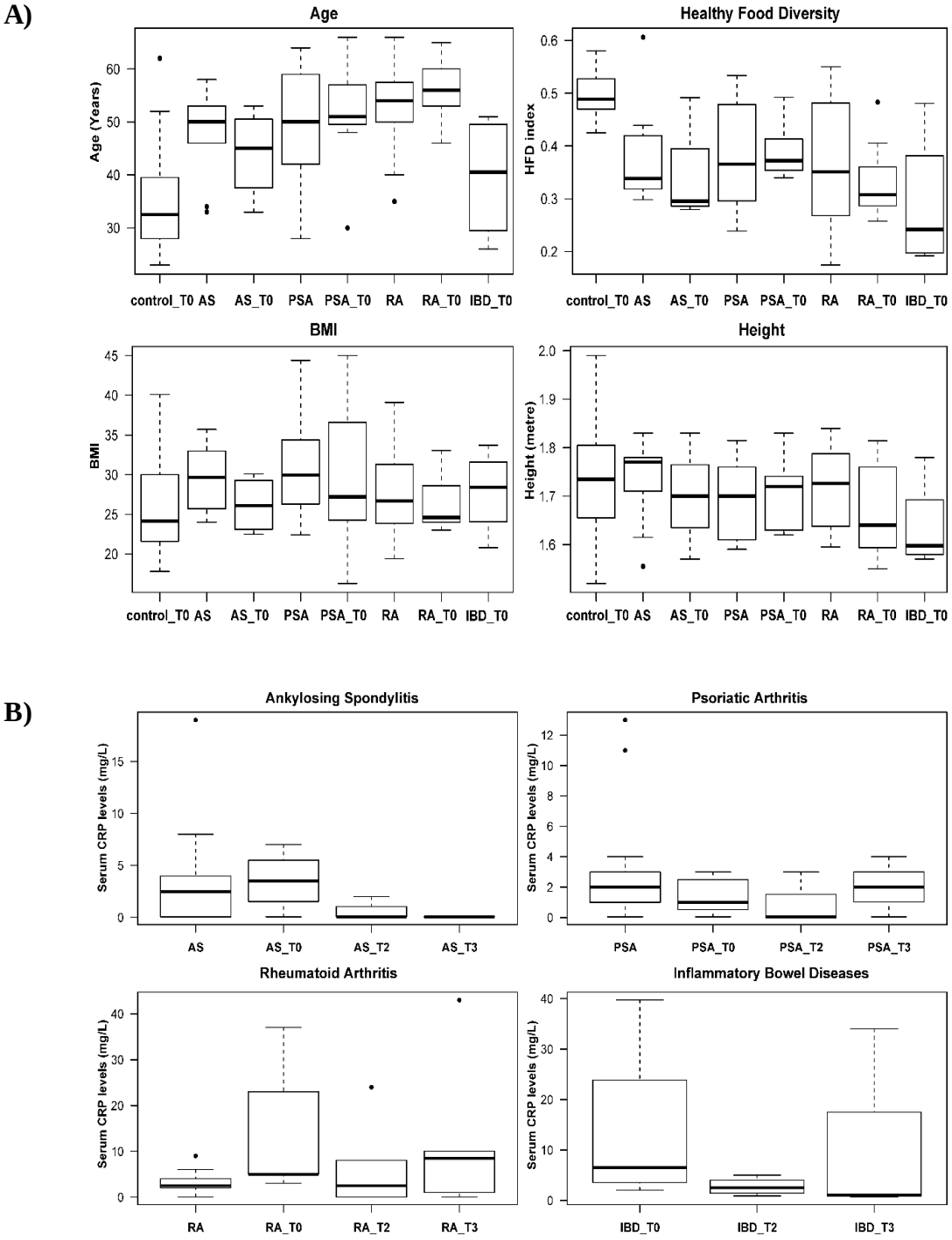


Figure 1 Significantly different meta-data characteristics of the disease populations compared to healthy controls

Meta-data observed to be significantly different between the controls and the disease samples as shown in Figure 1A. Figure 1B shows the significant differences in C-reactive protein (CRP) levels as observed in the different diseased groups. AS, PSA and RA represents established samples with long-term biologics treatment. T0 represents timepoint before the biologics treatment and T2 and T3 represents 3 months and 6 months after T0.

3.4.2 Microbiota characterisation

The MISeq mWGS sequencing run generated a total of 341.87 gigabases of quality filtered, contaminant removed sequence data comprising of 9,240,682±1,857,340 reads per sample. Mapping these reads to reference databases using the HUMAnN2 pipeline provided information on 348 bacterial species and 1,216,060 UniRef genefamilies resulting in 740 MetaCyc annotated pathways.

3.4.3 Taxonomic diversity is partially associated with Biologic treatment

The α-diversity for the control samples was stable over the timeseries **(Supplementary table 5, Supplementary table 6A)**. The IBD biologic naive samples (T0) showed a reduced number of observed species compared to control T0 and AS (T0) and RA (T0, T2 and T3) samples showed a higher number of observed species compared to control T0. This increase in richness was retained in the unrelated samples in the established RA group **(Supplementary table 6B, Figure 2)**. Other established arthritic diseases did not present with an altered richness compared to controls. IBD was also associated with a reduced gene count and Shannon index compared to controls **(Supplementary table 6B, Figure 2D,E)**. No significant alterations in α-diversity were observed for PSA in the treatment naive or established groups **(Supplementary table 6)**.

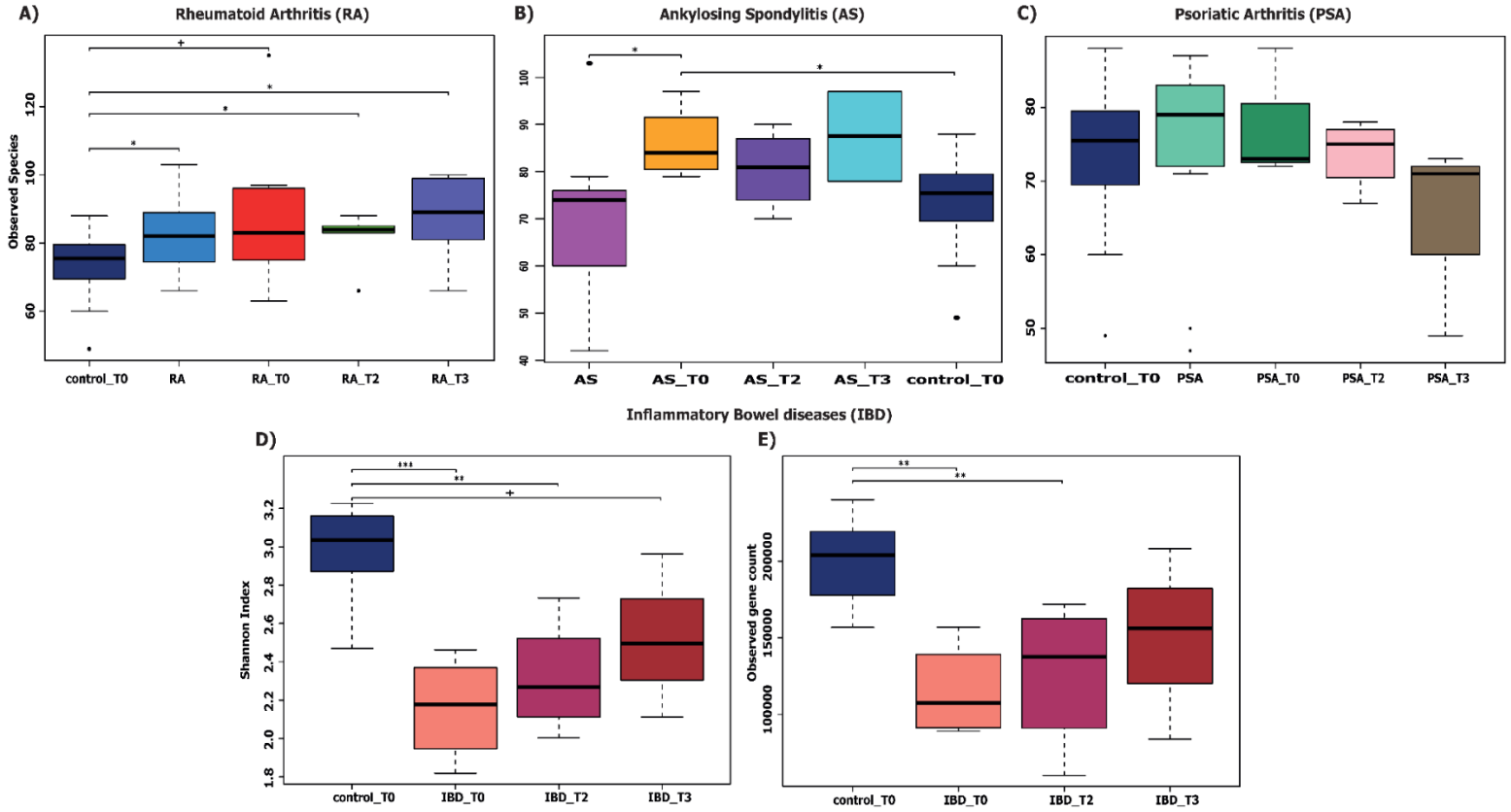


Figure 2 Significant within-sample taxonomic and functional diversity associated with various immune disorders

Differences in α-diversity was observed between the controls and various diseased samples. In plot A), B) and C) the number of unique species shows prominent differences in Rheumatoid arthritis, Ankylosing spondylitis but no significant difference in Psoriatic arthritis compared to controls. In plot D) Shannon index reflects significant within-sample variations in IBD against controls. Plot E) highlights significant difference in functional diversity as defined by gene count in IBD. Significance is defined by P-adjusted values. * ≤ 0.05, ** ≤ 0.005

3.4.4 Distinct β-diversity is observed with established and *de-novo* disease groups

The control timeseries samples showed a stable β-diversity with taxonomic datasets as no significant alterations were observed with BCA analysis on Spearman dissimilarity **(Supplementary table 7)**. The IBD and AS timeseries samples showed an altered β-diversity when compared against the control samples which was not observed with RA and PSA timeseries samples **(Supplementary table 7, Figure 3)**. Comparison of established disease groups against the controls identified no differences in β-diversity **(Supplementary table 7)**. The BCA ordination analysis showed that the established disease samples were closer to the controls than the timeseries samples in all cases **(Figure 3)**. β-diversity was different between all disease groups except between PSA and AS samples **(Supplementary table 7)**.

The Zhang *et al.* RA cohort showed a significant difference in β-diversity (Spearman distance) based on the group stratifications **(Supplementary table 7)**. Integrating the Zhang *et al.* RA taxonomic dataset with the current RA dataset revealed that the primary separation (X-axis) observed was between cohorts **(Supplementary table 7, Figure 4A)**. However, along the Y-axis, RA (T3) and established samples from our dataset moved away from RA (T0) and towards controls. This was also observed in Zhang *et al.* cohort where the treated samples were observed to be between control and naive RA samples. This suggests a potential restoration of the microbiota towards control-like following treatment.

The functional datasets reflected the results observed in the taxonomic datasets in the IBD samples which showed statistically significant separation where IBD (T3) samples were closer to controls compared to IBD (T0) treatment naive samples **(Supplementary table 8, Figure 5)**.

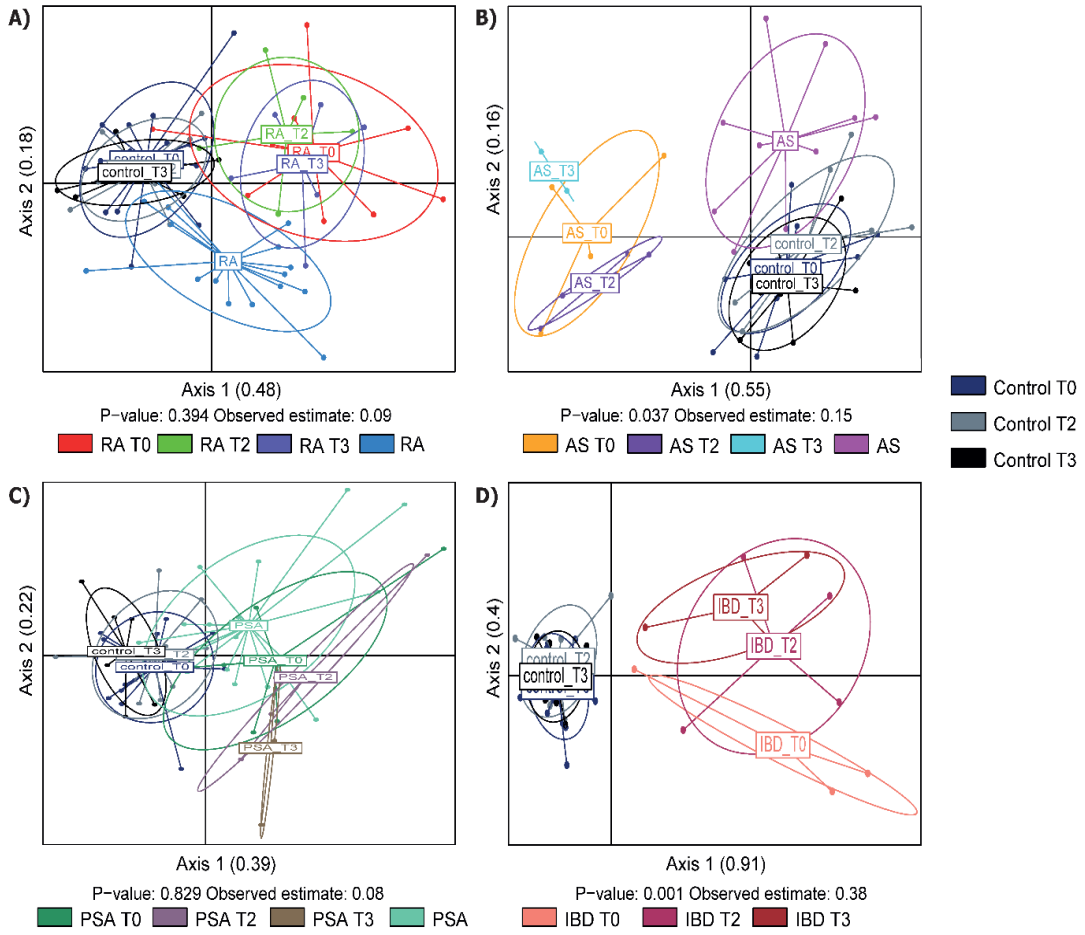


Figure 3 Global gut microbiota β-diversity in various immune disorders
The ordination represents between-class analysis (bca) using Spearman distance on taxonomic composition, stratified by groups and time-points. A) represents Rheumatoid arthritis (RA) B) represents Ankylosing spondylitis (AS), C) represents Psoriatic arthritis (PSA) and D) represents Inflammatory Bowel diseases (IBD) samples. Significant associations were observed only in AS and IBD with group plus time-point stratification. Established groups are represented by RA, AS, PSA and time-point samples are suffixed by T0, T2 and T3.

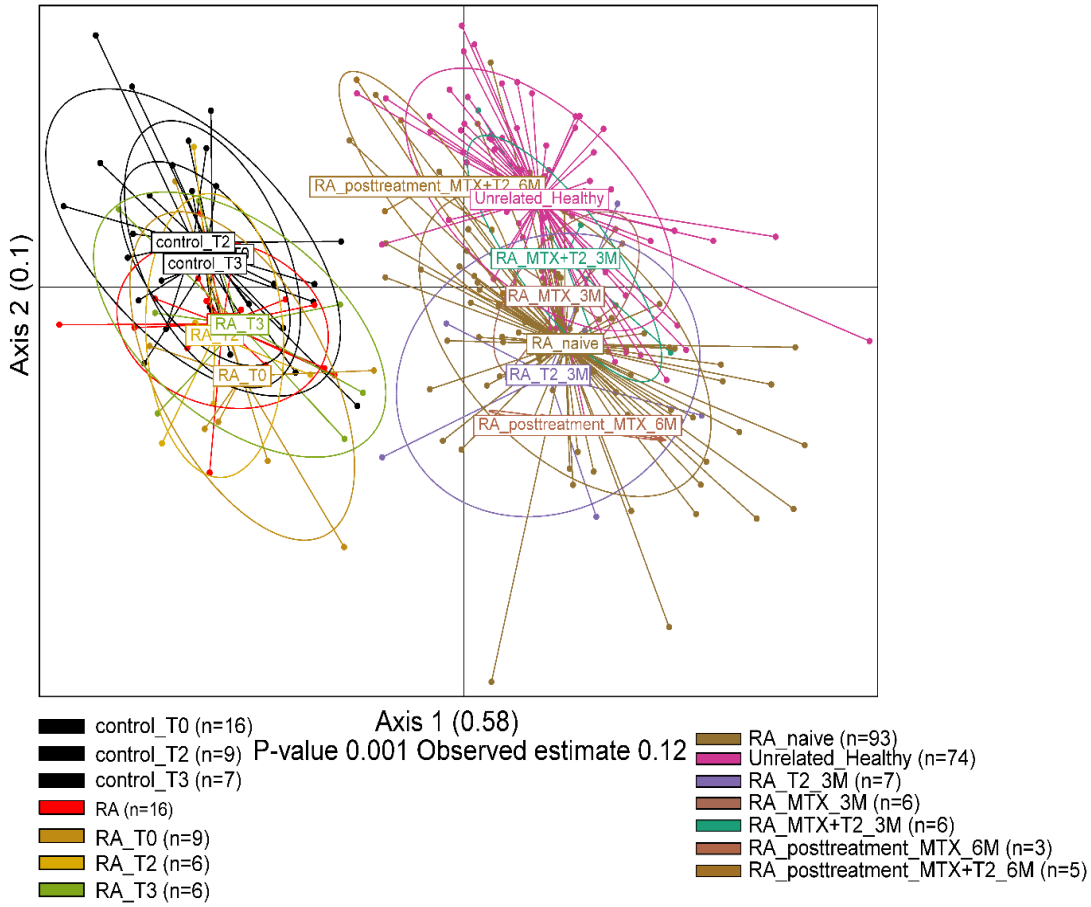


Figure 4 Comparison of RA microbiota profiles from this study with an independent RA dataset

Separation based on cohort is observed in both datasets. Axis 1 represents the primary split between the two rheumatoid arthritis cohorts. A trend is observed in both RA cohort showing movement away from treatment partly towards controls.

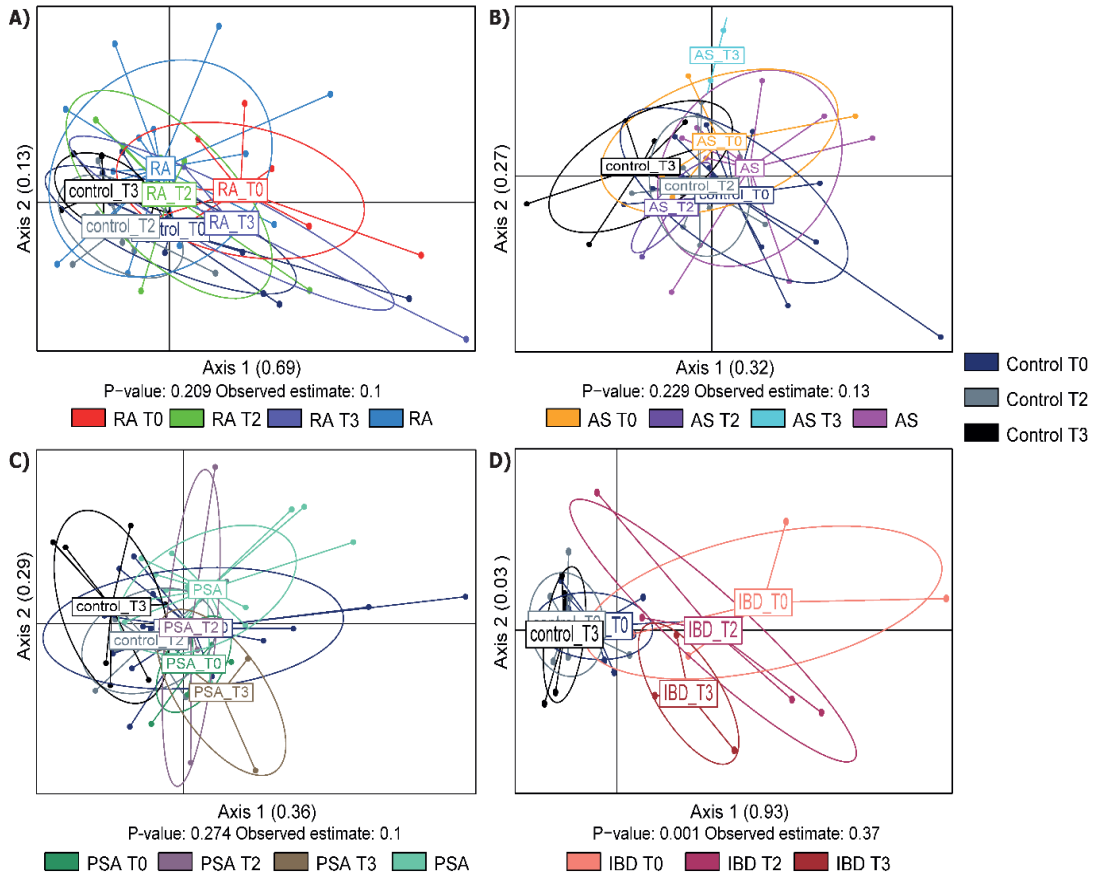


Figure 5 Functional β-diversity of the gut microbiota in various immune disorders

The ordination represents between-class analysis (bca) using Spearman distance on pathway abundance, stratified by groups and time-points. A) represents Rheumatoid arthritis (RA) B) represents Ankylosing spondylitis (AS), C) represents Psoriatic arthritis (PSA) and D) represents Inflammatory Bowel diseases (IBD) samples. Significant associations were observed only in IBD with group plus time-point stratification. Established groups are represented by RA, AS, PSA and time-point samples are suffixed by T0, T2 and T3.

The difference from T0 to T2 or T3 is an indicator of the degree of alterations in the β-diversity profiles based on taxonomic datasets across the timeseries **(Figure 6)**. Dissimilarity did not achieve statistical significance between T2-T0 in any disease groups against control **(Figure 6A)**. However, RA and IBD groups showed significant difference in dissimilarity in T3-T0 compared to controls **(Figure 6B)**.

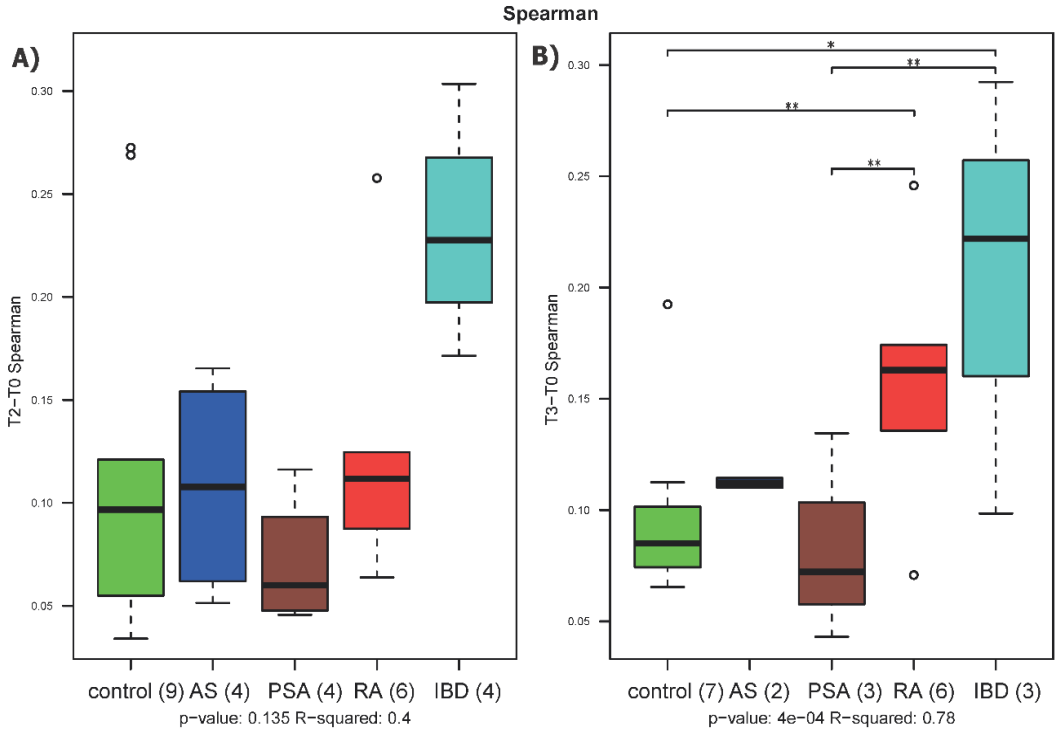


Figure 6 Increased dissimilarity between timepoints observed in treated disease states

Dissimilarity was defined as absolute distance as calculated by Spearman between subjects at T2 and T3 with T0 from different diseases. Plot A) highlights the distance observed for a subject between T2 and T0. Plot B) represents the distance observed between T3 and T0 for each subject. The subjects are coloured by Group category. Rheumatoid arthritis (RA), Ankylosing spondylitis (AS), Psoriatic arthritis (PSA) and Inflammatory Bowel diseases (IBD). Significance is defined by P-adjusted values. * ≤ 0.05, ** ≤ 0.005

3.4.5 Unique and shared taxonomic markers associated with disease groups

The analysis of abundant taxonomies identified various shared and unique bacterial species associated with different disease groups. Overall, 37 differentially abundant species, belonging to 23 different genera were shared between at least two disease groups. These included species belonging to known genera associated with potential immune modulation like *Alistipes, Bacteroides, Bifidobacterium, Clostridium, Collinsella, Desulfovibrio, Lactococcus, Paraprevotella, Rumminococcus* and *Streptococcus* **(Figure 7A-D)** (Liu et al., 2016;Clemente et al., 2018).

In the RA dataset, 23 species were differentially abundant between disease groups and controls **(Figure 7A, Supplementary table 9A)**. Amongst them, *Streptococcus gordonii, Streptococcus anginosus, Streptococcus infantis, Streptococcus mutans, Rothia dentocariosa, Collinsella intestinalis, Lactobacillus casei paracasei* and *Anaerotruncus colihominis* were prevalent only in the disease group. *Blautia hydrogenotrophica, Bacteroides salyeriae, Lactococcus lactis, Bacteroides clarus* were depleted in the disease samples. The established RA samples showed the presence of species from *Streptococcus, Collinsella* and *Veillonella* that were significantly associated with disease groups and *Bacteroides salyeriae, Blautia hydrogenotrophica, Lacococcus lactis* that were associated with control samples.

The AS dataset identified 43 different species belonging to 26 genera that were significantly differentially abundant between the groups from the pairwise comparisons **(Figure 7B, Supplementary table 9B)**. *Collinsella intestinalis, Clostridium hathewayi, Anaerotruncus colihominis, Clostridium symbiosum* and *Clostridium nexile* were identified in the AS timeseries samples and missing in the control and AS established samples. In the control group, *Lactococcus lactis, Bacteroides coprocola, Bacteroides stercoris Baceroides massiliensis, Bacteroides*

finegoldii, Desulfovibrio desulfuricans and *Bacteroides salyersiae* were more abundant. Certain taxa such as *Ruminococcus bromii, Bifidobacterium bifidum, Bacteroides stercoris, Bacteroides coprocola, Bacteroides massiliensis,* and *Eubacterium biforme* were observed across both established disease and control samples while they were depleted in the timeseries samples. The established AS disease group showed the presence of disease associated (*Anaerotruncus colihominis, Clostridium nexile*) and control associated taxa (*Desulfovibrio desulfuricans*, species of *Clostridium* and *Bacteroides*). *Eubacterium ventriosum* and *Bacteroides cellulosilyticus* were detected across all groups.

In the PSA dataset, 34 distinct species were identified to be significantly different between at least two pairwise comparisons **(Figure 7C, Supplementary table 9C)**. In the diseased timeseries samples, *Bacteroides faecis, Bacteroides finegoldii, Veillonella dispar* and *Lactococcus lactis* were distinctively depleted but were observed in established PSA and controls. Species such as *Roseburia intestinalis* and *Bacteroides fraigilis* were observed PSA T0 samples but were depleted in T2 and T3. *Roseburia intestinalis* was observed in both established PSA and control samples also. In the control samples, *Mitsuokella multacida* and *Bacteroides clarus* were prominent and reduced in all disease samples (established and timepoint).

Prominent group-specific signatures with 66 different species from 32 genera were seen in the IBD dataset. The differentially abundant species identified are well established with inflammation and IBD **(Figure 7D, Supplementary table 9D)**. In the disease group, *Blautia hansenii, Blautia producta, Lachnospiraceae bacterium species, Clostridiales bacterium species, Dorea, Clostridium nexile, Bacteroides fragilis, Enterococcus faecalis* and *Streptococcus gordonii* were dominant and completely absent in control group. *Ruminococcus gnavus* was observed in low

abundant in few control samples but was prominent in disease. Similarly, *Ruminococcus bromii* and *Bacteroides dorei* were observed in few IBD samples but were more abundant in controls. In the controls, *Roseburia hominis, Alistipes indistintus, Eubacterium rectale, Coprococcus species, Clostridium leptum, Bacteroides finegoldii, Bacteroides faecis, Paraprevotella species, Prevotella copri* to name a few were abundant and not detected in IBD samples.

A comparison of the common disease associated species across different disease types highlighted that the RA dataset shared the smallest number of differentially abundant species with other conditions whereas AS and IBD shared the most.

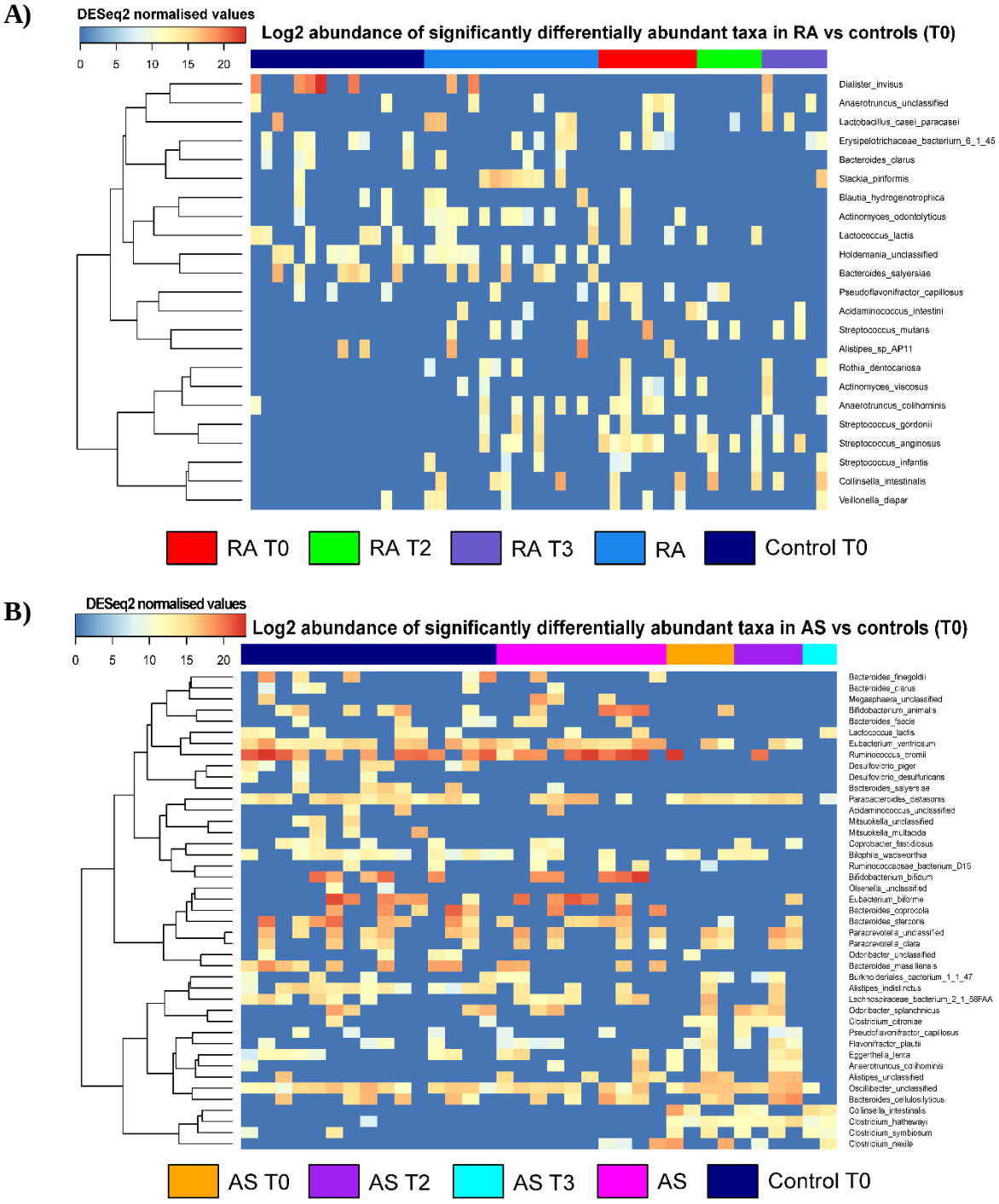


Figure 7 Heatmap of differentially abundant species in different diseases
Heatmap of the differentially abundant species between all pairwise comparisons identified by DESeq2 analysis in A) rheumatoid arthritis (RA) and B) Ankylosing spondylitis (AS). All pairwise comparisons were carried out within the disease samples and against control T0 samples. DESeq2 normalised values are plotted in the heatmap and the dendrogram represents Kendall distance based on the significant species.

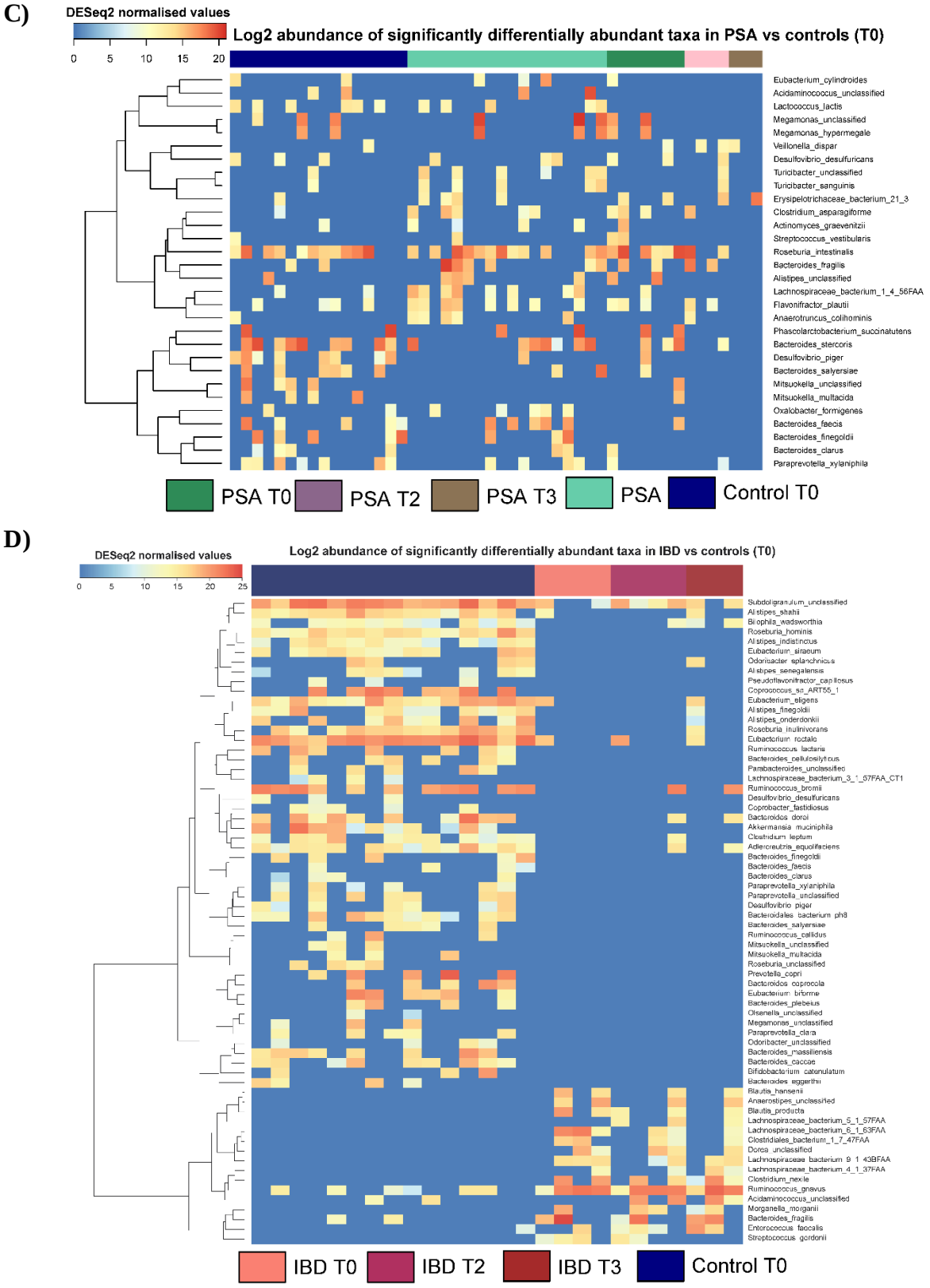


Figure 7 Heatmap of differentially abundant species in different diseases
Differentially abundant species between all pairwise comparisons identified by DESeq2 analysis in C) Psoriatic arthritis (PSA) and D) Inflammatory Bowel diseases (IBD). All pairwise comparisons were carried out within the disease samples and against control T0 samples. DESeq2 normalised values are plotted in the heatmap and the dendrogram represents Kendall distance based on the significant species.

3.4.6 Only IBD groups showed significantly different functional features

Investigation of the functional profiles determined by pathway abundances did not identify any significantly differentially abundant pathway in the arthritic groups. Pathways were observed to be significantly differentially abundant in the pairwise comparison between IBD timeseries samples vs controls (T0) **(Supplementary table 10)**. Amongst these, two pathways were guanosine and adenosine biosynthesis pathways, which are part of DNA replication. Arginine dependent acid resistance and asparagine degradation pathways was found to be significantly different in multi-group comparisons but did not pass the statistical threshold during the pairwise analysis. Only Arginine degradation and putrescine biosynthesis pathways were significantly different in their abundance between IBD timeseries and controls.

3.4.7 Nominally significant associations were observed between key meta-data and significant species

Key meta-data such as immune cells populations, CRP levels, diet (HFD index), BMI, weight, physical activity, and medical information such as adherence to medication and years since diagnosis were correlated with the identified differentially abundant species in different diseases. Significance of correlations observed between various meta-data and the species was defined by nominal p-value (≤ 0.05) **(Figure 8A-D)**.

In the RA dataset, different immune cells showed significant correlation with different species **(Figure 8A)**. *Blautia hydrogenotrophica* negatively correlated with white blood cells (WBCs) and neutrophils, whereas monocytes and basophils population positively correlated with *Streptococcus anginosus* and *Streptococcus infantis* respectively. CRP levels showed significant negative correlation with *Bacteroides clarus* and *Collinsella intestinalis*. Morisky score did not show any

significant correlation with any taxa, while unclassified *Holdemania* species showed positive association with disease duration.

Different species of *Clostridium* were significantly negatively correlated with disease duration in the AS dataset **(Figure 8B)**. Different immune cells population excluding lymphocytes were negatively correlated with different species of *Bacteroides*. Physical activity negatively correlated with *Bacteroides finegoldii, Paraprevotella* species, *Eubacterium ventriosum* and positively correlated with *Clostridium nexile* and *Collinsella intestinalis*.

Correlation between the meta-data and the species significantly different in the PSA dataset highlighted significant positive association between Basophils, Lymphocytes, WBCs and Neutrophils with *Streptococcus vestibularis, Flavonifractor plautii, Clostridium asparagiforme* and *Lachnospiraceae* bacterium **(Figure 8C)**. HFD and physical activity did not show any significant correlation with significant species observed in PSA. Disease duration was significantly positively correlated with *Turicibacter* species and negatively correlated with *Desulfovibrio piger* and *Clostridium asparagriforms*. CRP levels were also significantly positively correlated with *Flavonifractor plautii, Lactococcus lactis, Actinomyces graevenitzii* and negatively correlated with *Bacteroides corprophilus*.

As the IBD dataset had various species depleted in the diseased state, those species could not be correlated with various meta-data that were available only for the diseased samples **(Figure 8D)**. Overall, BMI showed significant negative correlation with *Clostridium leptum, Bacteroides dorei, Alistipes finegoldii, Bacteroides Massilliensis, Ordibacter species* and *Eubacterium retale*. Species belonging to *Roseburia, Alistipes, Eubacterium, Bacteroides, Akkermansia, Rumminococcus* and *Clostridium* were significantly positively associated with physical activity while

Streptococcus, Enterococcus and *Bacteroides fragilis* were negatively correlated. Diet also correlated positively with *Ruminococcus, Lachnospiraceae* bacterium, *Dorea* and showed negative correlation with *Streptococcus gordonii* and *Bacteroides fragilis*. Significant negative correlation was observed between immune cells (WBCs, Monocytes and Neutrophils) and species of *Bacteroides, Dorea, Ruminococcus gnavus, Lachnospiraceae* and positively correlated with *Eubacterium* and *Alistipes* species. *Bacteroides dorei* and *Ruminococcus bromii* negatively correlated with CRP levels.

Similar disease associated species from different diseases showed similar correlation with immune cells, diet, and physical activity. However, these correlations were nominally significant and many of the known association were non-significant after p-value adjustment.

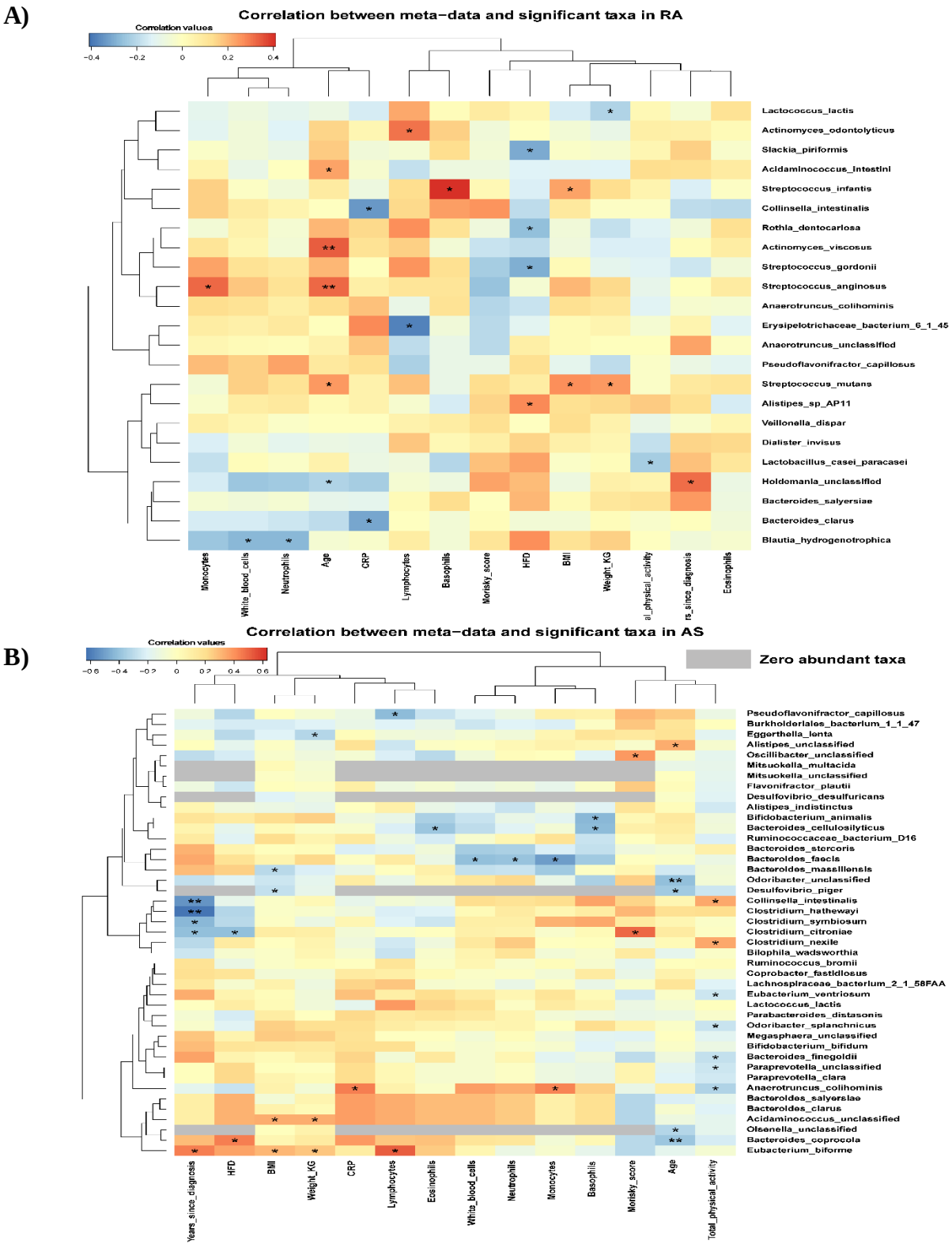


Figure 8 Correlation between significantly different species and clinical factors Heatmap of Kendall correlation between clinical meta-data and the abundances of significant taxa from Control T0 and the diseased samples per disease. Dendrogram is based on Euclidean distance on Kendall correlation. Plot A) represents correlation observed in RA dataset. B) shows correlation in AS dataset. Significance is based on nominal P-values. * ≤ 0.05, ** ≤ 0.005. Grey cells represent no correlation due to zero abundance of species in disease groups.

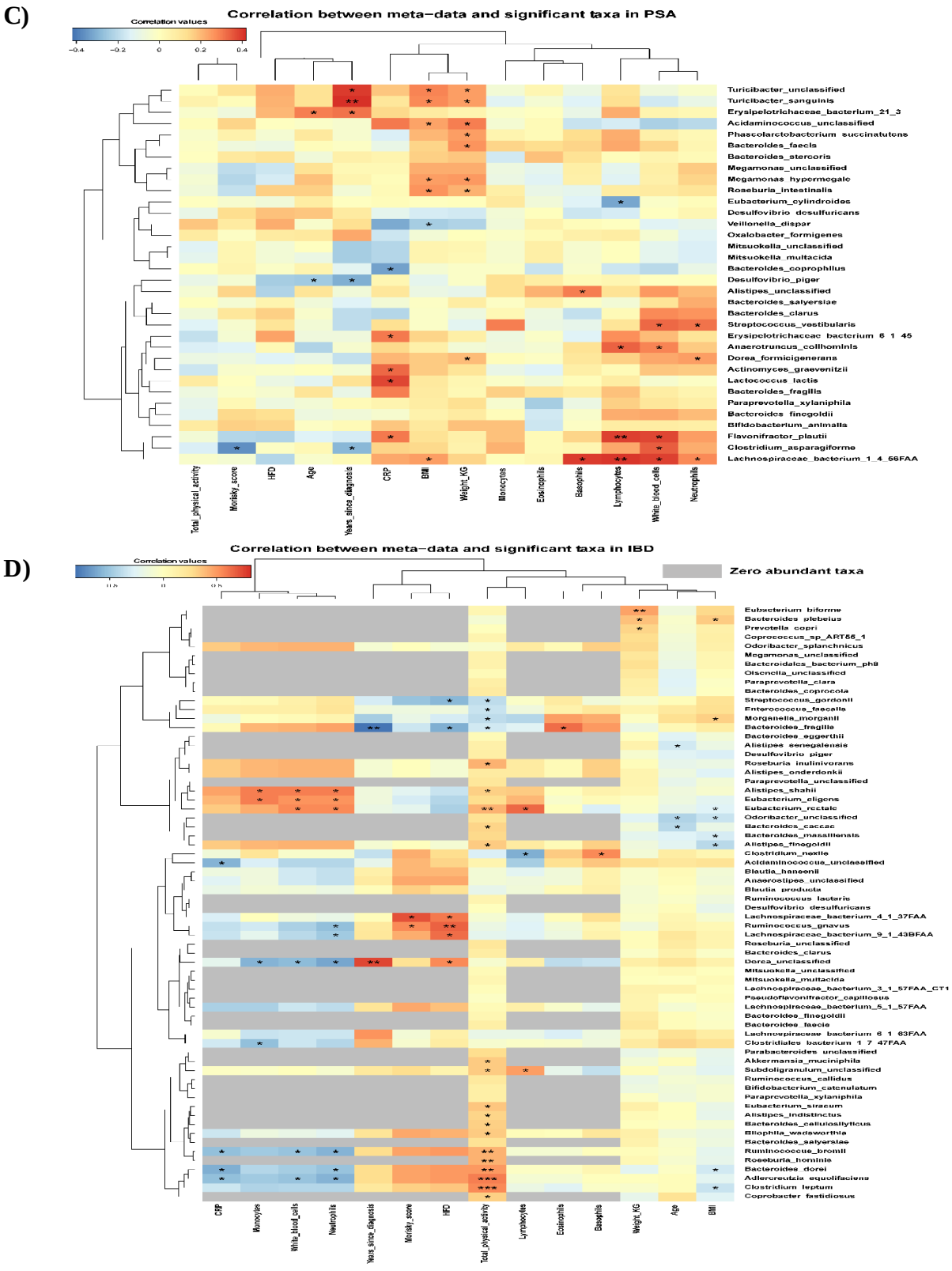


Figure 8 Correlation between significantly different species and clinical factors Heatmap of Kendall correlation between clinical meta-data and the abundances of significant taxa from Control T0 and the diseased samples per disease. Dendrogram is based on Euclidean distance on Kendall correlation. Plot C) represents correlation observed in PSA dataset. D) shows correlation in IBD dataset. Significance is based on nominal P-values. * ≤ 0.05, ** ≤ 0.005, *** ≤ 0.0005. Grey cells represent no correlation due to zero abundance of species in disease groups.

3.5 Discussion

This study investigated the dynamics of the gut microbiota populations associated with immunosuppressive biologic drugs. The study allowed for the analysis of α- and β-diversity of microbiome populations in disease states where we observe that biologics at least partially restore healthy-control associated taxa after long-term use of immunosuppressive drugs in two of the three arthritic diseases studied and in IBD. Extensive analysis of the clinical profile has highlighted some nominally significant associations with disease-associated taxa.

Diversity defined by within-sample variation and between sample dissimilarity showed a trend of movement of the gut microbiota populations towards the control subjects however, the dataset lacked the required statistical power. Despite the diversity being a key indicator of the stability of the gut microbiota population, previous cohort studies required large sample sizes to detect small alterations associated with the global microbiota composition (Zhang et al., 2015;Jackson et al., 2018), however similar to Zhang *et al* 2015, we were able to use timeseries to detect these treatment-associated alterations, and control for various inter-individual confounders. Previous studies have also noted that there was no observable association with alpha-diversity despite alterations in taxa-specific signatures (Zhang et al., 2015;Picchianti-Diamanti et al., 2018;Shapiro et al., 2019).

This is reflected in our study where we have identified taxonomic signature associated with the biologics treatment naive arthritic groups and noticeable difference from the biologics treated established groups, which partially resembled the control samples. Due to MTX treatment in the biologics naive groups, except for AS, the degree of restoration cannot be fully determined from a complete treatment naive status. The MTX treatment may be associated with partial restoration as highlighted

in two studies on RA by Zhang *et al.* and Picchianti-Diamanti *et al.* (Zhang et al., 2015;Picchianti-Diamanti et al., 2018). These two studies have also shown that treatment with Etanercept and glycosides of the traditional Chinese medicinal component *Tripterygium wilfordii* are also associated with changes in the gut microbiota towards healthy-like. Brazin *et al.* also observed that response to anti-TNF treatment was associated with alteration in the microbiota profile in Spondyloarthritis patients but not duration of treatment (Bazin et al., 2018).

The significantly differentially abundant taxa identified to be associated with different disease groups in this study or with the controls have also been evidenced in other independent studies (Clemente et al., 2018). This is indicative that despite the low sample size, the signature of the key markers is detectable. Different species of *Streptococcus* were observed to be more abundant in RA samples, which could be an indicator for molecular mimicry as *Streptococcal* antibodies have been linked with inflammation and autoimmunity (Cunningham, 2014). Similarly, *Rothia* is associated with infection in immunocompromised host and usually found in oral cavity (Ramanan et al., 2014). Its presence in our study could be allochthonous in nature and due to MTX medications. *Collinsella* is known to be associated with RA (Chen et al., 2016).

The AS disease group showed a dominance of *Clostridium* and *Collinsella* species. Different species of *Clostridium* are associated with immune homeostasis in the gut (Lopetuso et al., 2013). Interestingly taxa such as *Eubacterium* that are associated with anti-inflammatory properties in IBD were observed across established AS samples (Kanauchi et al., 2006).

The presence of species belonging to genus *Blautia, Lachnospiraceae, Clostridiales, Dorea* and *Bacteroides* in the disease subjects is interesting as these taxa

are observed to be depleted in other studies (Loh and Blaut, 2012;Sokol et al., 2017). This could potentially be a technical artefact, as these species were not observed in the controls. Enriched abundance of *Enterococcus faecalis, Ruminococcus gnavus* and *Streptococcus* species in IBD samples shows concordance with other studies (Liu et al., 1995;Zhou et al., 2016;Hall et al., 2017). On the other hand, taxa associated with healthy diet or healthy people such as *Roseburia, Eubacterium rectale, Ruminococcus bromii, Coprococcus* and *Bacteroides* were not found in IBD samples (Kang et al., 2010;Zhernakova et al., 2016).

PSA samples did not show prominent enriched taxa that were supported by literature. However, known taxa that are associated with healthy like certain species of *Bacteroides, Roseburia, Lactococccus lactis* were reduced in PSA (Kimoto et al., 2004;Tamanai-Shacoori et al., 2017). A recent study observed that treatment was not associated with any changes in the microbiota diversity or profiles in psoriatic patients (Shapiro et al., 2019).

The detection of different species of *Bacteroides* genus in both disease and control groups across various diseases is interesting as the responses from the immune system differs based on different species (Berezow et al., 2009). A recent study investigated the taxonomic profiles of different immune disorders and also observed increased *Streptococcus* abundance across all disease groups, and decreased abundance of *Corprococcus* and *Roseburia* in CD and RA, which was replicated in this study (Forbes et al., 2018). The prevalence of control associated taxa in the established groups of various diseases highlights restoration of gut microbiota towards normal-like.

The relationships between microbiota and various confounders are complex. Significant correlation between the significant taxa in various diseases and various

meta-data reflects established roles. This includes the relationship between immune cells and *Streptococcal* species (Guyre et al., 1990;Ellis et al., 2015), *Ruminococcus* species with inflammation as determined by immune cell population and CRP levels (Henke et al., 2019). Species from *Ruminococcus* genus is also linked with healthy diet (Ze et al., 2012). The positive correlation observed between physical activity and *Roseburia, Alistipes, Eubacterium* and *Akkermansia* is also documented in literature (Chen et al., 2018). These relationships are observed in parts across different diseases in this study highlighting the complex interactions of the gut-host axis. The identification of these signatures is an indicator of their biological significance in maintaining host health.

This study is a pilot investigation of the changes in gut microbiota dynamics due to biologics treatment. Despite uncovering key and established signature, the study is limited by its sample size. Other associated factors include a younger control samples compared to diseased groups, lack of representation of complete treatment naive disease cohort which could reflect the initial microbiota configuration at the onset or in early cases. The lack of serum profiles in the controls also limits the analysis to only disease samples without proper representation of those features in the general population. This study is an observational study and any association observed with the different disease groups or clinical factors does not imply any causation. Despite these limitations, the study highlights potential association of biologics and other immunosuppressants as supported by literature with the gut microbiota population and prevalence of different taxa.

In conclusion, we have highlighted that there are microbiota taxonomic associations related to treatment and remission status of the disease when treated with biologics, even if the subjects were previously treated with methotrexate. These

taxonomic alterations have the potential to influence microbiota-associated inflammation, remission status, effectiveness of therapy and may be important in maintaining remission.

3.6 Acknowledgements

Acknowledgements

We kindly acknowledge the academic support infrastructure of the APC Microbiome Ireland, University College Cork, Cork.

Competing Interests

None to declare

Funding

This work was partly supported by the Irish Centre for Arthritis Research and Education (ICARE). The authors are supported in part by Science Foundation Ireland (SFI) in the form of a research centre grant (SFI/12/RC/2273). IBJ and MD are funded by an SFI Starting Investigator Research Grant (SIRG) (13/SIRG/2128).

Ethics approval

Clinical Research Ethics Committee of the Cork Teaching Hospitals (CREC)

Contributors

MD carried out the bioinformatics analysis, compilation, and interpretation of the work, drafting and revising of the manuscript submitted. DMK, OC and MGM contributed to the concept, design of the work, acquisition of data, and reviewed the manuscript before submission. AM carried out DNA extraction and sequencing. SM, JB, JR, SH, MP, GM and CM contributed to the design of the work, acquisition of data and reviewed the manuscript before submission. PWOT contributed to the concept and design of the work, interpretation of the work, drafting and revising of the manuscript submitted. FS contributed to the concept and design of the work and revising of the manuscript submitted. IBJ contributed to the concept and design of the work, interpretation of the work, drafting and revising of the manuscript submitted.

3.7 References

1. Adlerberth, I., and Wold, A.E. (2009). Establishment of the gut microbiota in Western infants. *Acta Paediatr* 98**,** 229-238.
2. Arvikar, S.L., and Fisher, M.C. (2011). Inflammatory bowel disease associated arthropathy. *Curr Rev Musculoskelet Med* 4**,** 123-131.
3. Atzeni, F., Sarzi-Puttini, P., Botsios, C., Carletto, A., Cipriani, P., Favalli, E.G., Frati, E., Foschi, V., Gasparini, S., Giardina, A., Gremese, E., Iannone, F., Sebastiani, M., Ziglioli, T., Biasi, D., Ferri, C., Galeazzi, M., Gerli, R., Giacomelli, R., Gorla, R., Govoni, M., Lapadula, G., Marchesoni, A., Salaffi, F., Punzi, L., Triolo, G., and Ferraccioli, G. (2012). Long-term anti-TNF therapy and the risk of serious infections in a cohort of patients with rheumatoid arthritis: comparison of adalimumab, etanercept and infliximab in the GISEA registry. *Autoimmun Rev* 12**,** 225-229.
4. Baumgart, D.C., and Carding, S.R. (2007). Inflammatory bowel disease: cause and immunobiology. *Lancet* 369**,** 1627-1640.
5. Baumgart, D.C., and Sandborn, W.J. (2007). Inflammatory bowel disease: clinical aspects and established and evolving therapies. *Lancet* 369**,** 1641-1657.
6. Bazin, T., Hooks, K.B., Barnetche, T., Truchetet, M.E., Enaud, R., Richez, C., Dougados, M., Hubert, C., Barre, A., Nikolski, M., and Schaeverbeke, T. (2018). Microbiota Composition May Predict Anti-Tnf Alpha Response in Spondyloarthritis Patients: an Exploratory Study. *Sci Rep* 8**,** 5446.
7. Berezow, A.B., Ernst, R.K., Coats, S.R., Braham, P.H., Karimi-Naser, L.M., and Darveau, R.P. (2009). The structurally similar, penta-acylated lipopolysaccharides of Porphyromonas gingivalis and Bacteroides elicit strikingly different innate immune responses. *Microb Pathog* 47**,** 68-77.
8. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30**,** 2114-2120.
9. Brown, E.M., Kenny, D.J., and Xavier, R.J. (2019). Gut Microbiota Regulation of T Cells During Inflammation and Autoimmunity. *Annu Rev Immunol* 37**,** 599-624.
10. Chen, J., Guo, Y., Gui, Y., and Xu, D. (2018). Physical exercise, gut, gut microbiota, and atherosclerotic cardiovascular diseases. *Lipids Health Dis* 17**,** 17.
11. Chen, J., Wright, K., Davis, J.M., Jeraldo, P., Marietta, E.V., Murray, J., Nelson, H., Matteson, E.L., and Taneja, V. (2016). An expansion of rare lineage intestinal microbes characterizes rheumatoid arthritis. *Genome Med* 8**,** 43.
12. Chow, J., Tang, H., and Mazmanian, S.K. (2011). Pathobionts of the gastrointestinal microbiota and inflammatory disease. *Curr Opin Immunol* 23**,** 473-480.
13. Clemente, J.C., Manasson, J., and Scher, J.U. (2018). The role of the gut microbiome in systemic inflammatory disease. *BMJ* 360**,** j5145.
14. Cunningham, M.W. (2014). Rheumatic fever, autoimmunity, and molecular mimicry: the streptococcal connection. *Int Rev Immunol* 33**,** 314-329.
15. Dean, L.E., Jones, G.T., Macdonald, A.G., Downham, C., Sturrock, R.D., and Macfarlane, G.J. (2014). Global prevalence of ankylosing spondylitis. *Rheumatology (Oxford)* 53**,** 650-657.
16. Dinno, A. (2017). dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums. R package version 1.3.4.

17. Dray, S., and Dufour, A.B. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software* 22**,** 1-20.
18. Drescher, L.S., Thiele, S., and Mensink, G.B. (2007). A new index to measure healthy food diversity better reflects a healthy diet than traditional measures. *J Nutr* 137**,** 647-651.
19. Ellis, G.T., Davidson, S., Crotta, S., Branzk, N., Papayannopoulos, V., and Wack, A. (2015). TRAIL+ monocytes and monocyte-related cells cause lung damage and thereby increase susceptibility to influenza-Streptococcus pneumoniae coinfection. *EMBO Rep* 16**,** 1203-1218.
20. Forbes, J.D., Chen, C.Y., Knox, N.C., Marrie, R.A., El-Gabalawy, H., De Kievit, T., Alfa, M., Bernstein, C.N., and Van Domselaar, G. (2018). A comparative study of the gut microbiota in immune-mediated inflammatory diseases-does a common dysbiosis exist? *Microbiome* 6**,** 221.
21. Franzosa, E.A., Mciver, L.J., Rahnavard, G., Thompson, L.R., Schirmer, M., Weingart, G., Lipson, K.S., Knight, R., Caporaso, J.G., Segata, N., and Huttenhower, C. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. *Nat Methods* 15**,** 962-968.
22. Garrett, S., Jenkinson, T., Kennedy, L.G., Whitelock, H., Gaisford, P., and Calin, A. (1994). A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. *J Rheumatol* 21**,** 2286-2291.
23. Guyre, P.M., Campbell, A.S., Kniffin, W.D., and Fanger, M.W. (1990). Monocytes and polymorphonuclear neutrophils of patients with streptococcal pharyngitis express increased numbers of type I IgG Fc receptors. *J Clin Invest* 86**,** 1892-1896.
24. Hagstromer, M., Oja, P., and Sjostrom, M. (2006). The International Physical Activity Questionnaire (IPAQ): a study of concurrent and construct validity. *Public Health Nutr* 9**,** 755-762.
25. Hall, A.B., Yassour, M., Sauk, J., Garner, A., Jiang, X., Arthur, T., Lagoudas, G.K., Vatanen, T., Fornelos, N., Wilson, R., Bertha, M., Cohen, M., Garber, J., Khalili, H., Gevers, D., Ananthakrishnan, A.N., Kugathasan, S., Lander, E.S., Blainey, P., Vlamakis, H., Xavier, R.J., and Huttenhower, C. (2017). A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients. *Genome Med* 9**,** 103.
26. Harvey, R.F., and Bradshaw, J.M. (1980). A simple index of Crohn's-disease activity. *Lancet* 1**,** 514.
27. Henke, M.T., Kenny, D.J., Cassilly, C.D., Vlamakis, H., Xavier, R.J., and Clardy, J. (2019). Ruminococcus gnavus, a member of the human gut microbiome associated with Crohn's disease, produces an inflammatory polysaccharide. *Proc Natl Acad Sci U S A* 116**,** 12672-12677.
28. Jackson, M.A., Verdi, S., Maxan, M.E., Shin, C.M., Zierer, J., Bowyer, R.C.E., Martin, T., Williams, F.M.K., Menni, C., Bell, J.T., Spector, T.D., and Steves, C.J. (2018). Gut microbiota associations with common diseases and prescription medications in a population-based cohort. *Nat Commun* 9**,** 2655.
29. Jeffery, I.B., Lynch, D.B., and O'toole, P.W. (2016). Composition and temporal stability of the gut microbiota in older persons. *ISME J* 10**,** 170-182.
30. Kanauchi, O., Fukuda, M., Matsumoto, Y., Ishii, S., Ozawa, T., Shimizu, M., Mitsuyama, K., and Andoh, A. (2006). Eubacterium limosum ameliorates experimental colitis and metabolite of microbe attenuates colonic inflammatory action with increase of mucosal integrity. *World J Gastroenterol* 12**,** 1071-1077.

31. Kang, S., Denman, S.E., Morrison, M., Yu, Z., Dore, J., Leclerc, M., and Mcsweeney, C.S. (2010). Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. *Inflamm Bowel Dis* 16**,** 2034-2042.
32. Kho, Z.Y., and Lal, S.K. (2018). The Human Gut Microbiome - A Potential Controller of Wellness and Disease. *Front Microbiol* 9**,** 1835.
33. Khor, B., Gardet, A., and Xavier, R.J. (2011). Genetics and pathogenesis of inflammatory bowel disease. *Nature* 474**,** 307-317.
34. Kimoto, H., Mizumachi, K., Okamoto, T., and Kurisaki, J. (2004). New Lactococcus strain with immunomodulatory activity: enhancement of Th1-type immune response. *Microbiol Immunol* 48**,** 75-82.
35. Landi, M., Maldonado-Ficco, H., Perez-Alamino, R., Maldonado-Cocco, J.A., Citera, G., Arturi, P., Sampaio-Barros, P.D., Flores Alvarado, D.E., Burgos-Vargas, R., Santos, E., Palleiro, D., Gutierrez, M.A., Vieira-Sousa, E., Pimentel-Santos, F., Paira, S.O., Berman, A., Barrezueta, C.V., Vazquez-Mellado, J., Collantes-Estevez, E., and Morteo, R.G.F.R.a.D.O.G. (2016). Gender differences among patients with primary ankylosing spondylitis and spondylitis associated with psoriasis and inflammatory bowel disease in an iberoamerican spondyloarthritis cohort. *Medicine (Baltimore)* 95**,** e5652.
36. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9**,** 357-359.
37. Ledingham, J., Snowden, N., and Ide, Z. (2017). Diagnosis and early management of inflammatory arthritis. *BMJ* 358**,** j3248.
38. Liu, J.T., Yeh, H.M., Liu, S.Y., and Chen, K.T. (2014). Psoriatic arthritis: Epidemiology, diagnosis, and treatment. *World J Orthop* 5**,** 537-543.
39. Liu, X., Zeng, B., Zhang, J., Li, W., Mou, F., Wang, H., Zou, Q., Zhong, B., Wu, L., Wei, H., and Fang, Y. (2016). Role of the Gut Microbiome in Modulating Arthritis Progression in Mice. *Sci Rep* 6**,** 30594.
40. Liu, Y., Van Kruiningen, H.J., West, A.B., Cartun, R.W., Cortot, A., and Colombel, J.F. (1995). Immunocytochemical evidence of Listeria, Escherichia coli, and Streptococcus antigens in Crohn's disease. *Gastroenterology* 108**,** 1396-1404.
41. Loh, G., and Blaut, M. (2012). Role of commensal gut bacteria in inflammatory bowel diseases. *Gut Microbes* 3**,** 544-555.
42. Lopetuso, L.R., Scaldaferri, F., Petito, V., and Gasbarrini, A. (2013). Commensal Clostridia: leading players in the maintenance of gut homeostasis. *Gut Pathog* 5**,** 23.
43. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15**,** 550.
44. Maier, L., Pruteanu, M., Kuhn, M., Zeller, G., Telzerow, A., Anderson, E.E., Brochado, A.R., Fernandez, K.C., Dose, H., Mori, H., Patil, K.R., Bork, P., and Typas, A. (2018). Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature* 555**,** 623-628.
45. Mckeown, N.M., Day, N.E., Welch, A.A., Runswick, S.A., Luben, R.N., Mulligan, A.A., Mctaggart, A., and Bingham, S.A. (2001). Use of biological markers to validate self-reported dietary intake in a random sample of the European Prospective Investigation into Cancer United Kingdom Norfolk cohort. *Am J Clin Nutr* 74**,** 188-196.
46. Mezoff, E.A., and Dykes, D.M. (2015). The Changing Face of Very Early-Onset Inflammatory Bowel Disease. *J Pediatr* 167**,** 508-509.

47. Morisky, D.E., Green, L.W., and Levine, D.M. (1986). Concurrent and predictive validity of a self-reported measure of medication adherence. *Med Care* 24**,** 67-74.
48. Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., Minchin, P.R., O'hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., and Wagner, H. (2017). vegan: Community Ecology Package. R package version 2.4-3.
49. Pasolli, E., Schiffer, L., Manghi, P., Renson, A., Obenchain, V., Truong, D.T., Beghini, F., Malik, F., Ramos, M., Dowd, J.B., Huttenhower, C., Morgan, M., Segata, N., and Waldron, L. (2017). Accessible, curated metagenomic data through ExperimentHub. *Nat Methods* 14**,** 1023-1024.
50. Peloquin, J.M., and Nguyen, D.D. (2013). The microbiota and inflammatory bowel disease: insights from animal models. *Anaerobe* 24**,** 102-106.
51. Picchianti-Diamanti, A., Panebianco, C., Salemi, S., Sorgi, M.L., Di Rosa, R., Tropea, A., Sgrulletti, M., Salerno, G., Terracciano, F., D'amelio, R., Lagana, B., and Pazienza, V. (2018). Analysis of Gut Microbiota in Rheumatoid Arthritis Patients: Disease-Related Dysbiosis and Modifications Induced by Etanercept. *Int J Mol Sci* 19.
52. Powell-Tuck, J., Bown, R.L., and Lennard-Jones, J.E. (1978). A comparison of oral prednisolone given as single or multiple daily doses for active proctocolitis. *Scand J Gastroenterol* 13**,** 833-837.
53. Prevoo, M.L., Van 'T Hof, M.A., Kuper, H.H., Van Leeuwen, M.A., Van De Putte, L.B., and Van Riel, P.L. (1995). Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum* 38**,** 44-48.
54. Ramanan, P., Barreto, J.N., Osmon, D.R., and Tosh, P.K. (2014). Rothia bacteremia: a 10-year experience at Mayo Clinic, Rochester, Minnesota. *J Clin Microbiol* 52**,** 3184-3189.
55. Rogier, R., Evans-Marin, H., Manasson, J., Van Der Kraan, P.M., Walgreen, B., Helsen, M.M., Van Den Bersselaar, L.A., Van De Loo, F.A., Van Lent, P.L., Abramson, S.B., Van Den Berg, W.B., Koenders, M.I., Scher, J.U., and Abdollahi-Roodsaz, S. (2017). Alteration of the intestinal microbiome characterizes preclinical inflammatory arthritis in mice and its modulation attenuates established arthritis. *Sci Rep* 7**,** 15613.
56. Schoels, M., Aletaha, D., Funovits, J., Kavanaugh, A., Baker, D., and Smolen, J.S. (2010). Application of the DAREA/DAPSA score for assessment of disease activity in psoriatic arthritis. *Ann Rheum Dis* 69**,** 1441-1447.
57. Shapiro, J., Cohen, N.A., Shalev, V., Uzan, A., Koren, O., and Maharshak, N. (2019). Psoriatic patients have a distinct structural and functional fecal microbiota compared with controls. *J Dermatol* 46**,** 595-603.
58. Simone, D., Al Mossawi, M.H., and Bowness, P. (2018). Progress in our understanding of the pathogenesis of ankylosing spondylitis. *Rheumatology (Oxford)* 57**,** vi4-vi9.
59. Smolen, J.S., Aletaha, D., Barton, A., Burmester, G.R., Emery, P., Firestein, G.S., Kavanaugh, A., Mcinnes, I.B., Solomon, D.H., Strand, V., and Yamamoto, K. (2018). Rheumatoid arthritis. *Nat Rev Dis Primers* 4**,** 18001.
60. Sokol, H., Leducq, V., Aschard, H., Pham, H.P., Jegou, S., Landman, C., Cohen, D., Liguori, G., Bourrier, A., Nion-Larmurier, I., Cosnes, J., Seksik, P.,

Langella, P., Skurnik, D., Richard, M.L., and Beaugerie, L. (2017). Fungal microbiota dysbiosis in IBD. *Gut* 66**,** 1039-1048.
61. Southwood, T.R., Petty, R.E., Malleson, P.N., Delgado, E.A., Hunt, D.W., Wood, B., and Schroeder, M.L. (1989). Psoriatic arthritis in children. *Arthritis Rheum* 32**,** 1007-1013.
62. Tamanai-Shacoori, Z., Smida, I., Bousarghin, L., Loreal, O., Meuric, V., Fong, S.B., Bonnaure-Mallet, M., and Jolivet-Gougeon, A. (2017). Roseburia spp.: a marker of health? *Future Microbiol* 12**,** 157-170.
63. Team, R.C. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing. *Vienna, Austria*.
64. Truong, D.T., Franzosa, E.A., Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., and Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods* 12**,** 902-903.
65. Veale, D.J., and Fearon, U. (2018). The pathogenesis of psoriatic arthritis. *Lancet* 391**,** 2273-2284.
66. Yu, Z., and Morrison, M. (2004). Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* 36**,** 808-812.
67. Ze, X., Duncan, S.H., Louis, P., and Flint, H.J. (2012). Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. *ISME J* 6**,** 1535-1543.
68. Zhang, X., Zhang, D., Jia, H., Feng, Q., Wang, D., Liang, D., Wu, X., Li, J., Tang, L., Li, Y., Lan, Z., Chen, B., Li, Y., Zhong, H., Xie, H., Jie, Z., Chen, W., Tang, S., Xu, X., Wang, X., Cai, X., Liu, S., Xia, Y., Li, J., Qiao, X., Al-Aama, J.Y., Chen, H., Wang, L., Wu, Q.J., Zhang, F., Zheng, W., Li, Y., Zhang, M., Luo, G., Xue, W., Xiao, L., Li, J., Chen, W., Xu, X., Yin, Y., Yang, H., Wang, J., Kristiansen, K., Liu, L., Li, T., Huang, Q., Li, Y., and Wang, J. (2015). The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nat Med* 21**,** 895-905.
69. Zhernakova, A., Kurilshikov, A., Bonder, M.J., Tigchelaar, E.F., Schirmer, M., Vatanen, T., Mujagic, Z., Vila, A.V., Falony, G., Vieira-Silva, S., Wang, J., Imhann, F., Brandsma, E., Jankipersadsing, S.A., Joossens, M., Cenit, M.C., Deelen, P., Swertz, M.A., Lifelines Cohort, S., Weersma, R.K., Feskens, E.J., Netea, M.G., Gevers, D., Jonkers, D., Franke, L., Aulchenko, Y.S., Huttenhower, C., Raes, J., Hofker, M.H., Xavier, R.J., Wijmenga, C., and Fu, J. (2016). Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* 352**,** 565-569.
70. Zhou, Y., Chen, H., He, H., Du, Y., Hu, J., Li, Y., Li, Y., Zhou, Y., Wang, H., Chen, Y., and Nie, Y. (2016). Increased Enterococcus faecalis infection is associated with clinically active Crohn disease. *Medicine (Baltimore)* 95**,** e5019.

3.8 Supplementary Tables

Supplementary table 1 Sample characteristics of the treatment naive subjects

A) Sample count, mean and standards deviation of different meta-data variables observed in the controls and time-point samples

Groups	control			AS			PSA			RA			IBD		
Time point	**T0**	**T2**	**T3**	**T0**	**T2**	**T3**	**T0**	**T2**	**T3**	**T0**	**T2**	**T3**	**T0**	**T2**	**T3**
Sample size	16	9	7	4	3	2	7	4	3	9	6	6	4	4	3
Sex(Male/Female)	9/7	5/4	3/4	2/2	2/1	2/0	3/4	2/4	2/1	5/4	3/3	3/3	1/3	1/3	1/2
Smoking (Yes/No/Previous)	1/11/4	0/7/2	0/7/0	0/3/1	0/2/1	0/1/1	3/4/0	1/3/0	1/2/0	3/3/3	2/2/2	2/2/2	0/3/1	0/3/1	0/2/1
Alcohol (Yes/No/Previous)	16/0/0	9/0/0	7/0/0	2/2/0	1/2/0	1/1/0	5/2/0	3/1/0	2/1/0	8/1/0	5/1/0	5/1/0	4/0/0	4/0/0	3/0/0
Abdominal Surgery (Yes/No)	NA	NA	NA	3/1	2/1	1/1	2/5	1/3	1/2	7/2	5/1	5/1	3/1	3/1	2/1
History of immune disorders (Yes/No)	NA	NA	NA	3/1	2/1	1/1	4/2	3/0	3/0	4/5	2/4	2/4	3/0	3/0	2/0
Age	35.44 ± 11.22	36.89 ± 9.97	38.71 ± 10.8	44 ± 8.6	48 ± 6	45 ± 4.24	51.43 ± 11.21	49.25 ± 14.75	48.67 ± 18.01	55.67 ± 6.98	57.17 ± 7.44	57.67 ± 7.31	39.5 ± 11.96	40.25 ± 12.34	36.67 ± 11.24
BMI	25.54 ± 5.7	26.61 ± 6.72	27.86 ± 8.05	26.2 ± 3.67	28.1 ± 3.57	26.85 ± 4.45	30.03 ± 9.97	25.12 ± 7.67	21.87 ± 5.16	26.36 ± 3.7	28.03 ± 3.92	27.63 ± 3.36	27.83 ± 5.38	29.4 ± 3.21	28.4 ± 2.48
Height (m)	1.74 ± 0.12	1.73 ± 0.13	1.7 ± 0.13	1.70 ± 0.11	1.70 ± 0.13	1.76 ± 0.09	1.70 ± 0.08	1.74 ± 0.08	1.77 ± 0.06	1.67 ± 0.09	1.68 ± 0.10	1.68 ± 0.10	1.64 ± 0.10	1.64 ± 0.10	1.65 ± 0.12
Weight (KG)	76.83 ± 16.53	78.46 ± 17.32	80.46 ± 21.11	75.85 ± 16.05	80.73 ± 18.6	84.45 ± 22.7	85.07 ± 22.04	75.08 ± 19.35	67.8 ± 17.66	74.4 ± 17.07	80.07 ± 18.93	78.82 ± 17.45	73.53 ± 8.3	78.35 ± 5.36	77.13 ± 3.54
Total physical activity	4455.94 ± 4137.6	5327.17 ± 4011.62	6169.83 ± 6374.05	10875.38 ± 10181.4	16424 ± 14126.85	28434 ± 17768.18	12101.07 ± 15759.18	3759 ± 3123.53	7531 ± 1975.51	12780 ± 24378.86	4831.08 ± 4818.16	4305.58 ± 5789.11	372.67 ± 189.13	3270 ± 4951.75	1006.67 ± 1036.38
HFD	0.5 ± 0.05	0.49 ± 0.04	0.52 ± 0.03	0.34 ± 0.1	0.31 ± 0.05	0.39 ± 0.02	0.39 ± 0.05	0.43 ± 0.07	0.41 ± 0.07	0.33 ± 0.07	0.33 ± 0.07	0.37 ± 0.08	0.29 ± 0.13	0.31 ± 0.13	0.35 ± 0.13
Disease activity	NA	NA	NA	Only performed on one sample across time point			22.04 ± 11.5	19.75 ± 16.44	9 ± 4.36	4.91 ± 0.65	4.24 ± 0.86	4.08 ± 1.08	5 ± 4.32	2.67 ± 2.31	0.5 ± 0.71

The number of samples per time-point in different disease groups for various meta-data variables. For immune disorder history, deviation from the total number represents lack of complete reports from all the subjects. Continuous information are represented by Mean and standard deviation (SD). Mean±SD. HFD, Healthy Food Diversity.

B) Serum profile, medication, and disease activity of the samples in different disease groups

Groups	AS			PSA			RA			IBD		
Time-point	**T0**	**T2**	**T3**	**T0**	**T2**	**T3**	**T0**	**T2**	**T3**	**T0**	**T2**	**T3**
MTX treatments (Yes/No)	0/4	0/3	0/2	6/1	4/0	3/0	9/0	6/0	6/0	2/2	1/3	1/2
NSAIDs (Yes/No)	4/0	3/0	2/0	6/1	4/0	3/0	9/0	6/0	6/0	NA	NA	NA
Morisky score	6.94 ± 0.92	8 ± 0	7.5 ± 0.71	6.5 ± 0.9	6.44 ± 1.94	7 ± 1.32	5.97 ± 1.02	7.17 ± 1.17	8 ± 0	5.38 ± 2.29	7.19 ± 0.55	7.67 ± 0.58
White blood cells	6.92 ± 2.22	5.63 ± 0.84	5.6 ± 0.99	8.69 ± 4.96	6 ± 1.93	6.77 ± 1.9	9.84 ± 2.29	7.4 ± 2.82	7.9 ± 2.99	8.72 ± 2.19	7.2 ± 2.82	10.17 ± 4.05
Neutrophils	4.75 ± 1.92	2.97 ± 0.55	2.9 ± 0.99	5.95 ± 4.82	3.21 ± 1.31	4.29 ± 2.44	7.2 ± 2.05	4.35 ± 2.59	4.85 ± 2.86	5.81 ± 1.63	4.3 ± 1.97	6.99 ± 3.78
Lymphocytes	1.48 ± 0.26	1.93 ± 0.23	1.95 ± 0.07	1.92 ± 0.72	1.98 ± 0.66	1.66 ± 0.39	1.61 ± 0.51	3.9 ± 4.52	2.05 ± 0.54	1.84 ± 0.6	1.98 ± 0.9	1.79 ± 0.45
Monocytes	0.45 ± 0.19	0.33 ± 0.06	0.4 ± 0.14	0.48 ± 0.16	0.44 ± 0.12	0.49 ± 0.05	0.69 ± 0.12	0.5 ± 0.15	0.48 ± 0.18	0.59 ± 0.13	0.43 ± 0.09	0.59 ± 0.11
Eosinophils	0.11 ± 0.07	0.19 ± 0.08	0.18 ± 0.13	0.17 ± 0.07	0.16 ± 0.05	0.17 ± 0.11	0.25 ± 0.22	16.63 ± 39.86	0.37 ± 0.22	0.42 ± 0.47	0.37 ± 0.18	0.64 ± 0.4
Basophils	0.04 ± 0.04	0.06 ± 0.01	0.06 ± 0.02	0.03 ± 0.03	0.02 ± 0.02	0.05 ± 0.05	0.05 ± 0.05	5.52 ± 13.46	0.03 ± 0.04	0.04 ± 0.04	0.04 ± 0.05	0.09 ± 0.01
CRP	3.51 ± 2.87	0.7 ± 1.13	0.05 ± 0	1.44 ± 1.25	1.03 ± 1.7	2.02 ± 1.98	13.71 ± 13.54	6.18 ± 9.25	11.84 ± 15.86	13.7 ± 17.51	2.73 ± 1.74	11.9 ± 19.14
Disease activity	Only performed on one sample across time point			22.04 ± 11.5	19.75 ± 16.44	9 ± 4.36	4.91 ± 0.65	4.24 ± 0.86	4.08 ± 1.08	5 ± 4.32	2.67 ± 2.31	0.5 ± 0.71

Observed serum profiles, medication and disease activity for all the time-point diseased samples Disease activity score for different disease is defined by (AS: BADSAI; PSA: DAPSA; RA: DAS28; IBD: Harvey Bradshaw Index Score for Crohn's & Powell Tuck Score for UC). CRP, C-Reactive protein.

Supplementary table 2 Sample characteristics of Established disease samples

	AS	PSA	RA	P-adjusted AS - control_ T0	P-adjusted control_ T0 - PSA	P-adjusted control_ T0 - RA
Sample size	10	18	16			
Sex(Male/Female)	9/1	7/11	9/7			
Smoking (Yes/No/Previous)	3/4/3	4/9/5	2/6/8			
Alcohol (Yes/No/Previous)	8/2/0	13/5/0	12/4/0			
Abdominal Surgery (Yes/No)	0/10	4/2	7/0			
History of immune disorders (Yes/No)	6/3	8/9	11/5			
MTX treatments (Yes/No)	0/7	6/12	6/4			
NSAIDs (Yes/No)	9/1	16/2	13/3			
Age	47.6 ± 8.15	49.44 ± 11.07	52.81 ± 7.51	**2.1E-02**	**6.3E-03**	**8.1E-04**
BMI	29.34 ± 4.11	31.06 ± 6.71	27.79 ± 5.54	5.2E-02	1.2E-02	2.4E-01
Height (m)	173.48 ± 8.73	159.24 ± 39.99	171.74 ± 8.21	**1.6E-04**	2.2E-05	**9.2E-06**
Weight (KG)	88.65 ± 15.13	88.42 ± 20.14	82.2 ± 17.72	1.2E-01	1.8E-01	4.9E-01
Total physical activity	12763.17 ± 15413.33	3968.28 ± 3105.56	10248.19 ± 10119.35	4.9E-01	8.8E-01	2.4E-01
HFD	0.38 ± 0.09	0.38 ± 0.1	0.37 ± 0.12	**1.7E-03**	**1.7E-03**	**3.9E-03**
Morisky_score	6.42 ± 1.31	6.61 ± 1.6	6.92 ± 1.54			
White_blood_cells	6.13 ± 1.86	7.25 ± 3.15	6.75 ± 2.16			
Neutrophils	3.23 ± 1.27	4.5 ± 3	4.1 ± 1.79			
Lymphocytes	2.14 ± 0.72	2.02 ± 0.78	1.87 ± 0.47			
Monocytes	0.34 ± 0.11	0.41 ± 0.17	0.41 ± 0.2			
Eosinophils	0.23 ± 0.11	0.18 ± 0.11	0.22 ± 0.09			
Basophils	0.04 ± 0.02	0.04 ± 0.02	0.03 ± 0.01			
CRP	3.92 ± 5.86	3.48 ± 4.38	3.17 ± 2.44			

In the established samples, reported values are based on group wise stratification, mean ± standard deviation for various meta-data, including serum profiles, dietary, anthropometric measures, medication and medical history. Pairwise comparison with Dunn's test with adjustment is reported. Significance is defined by P-adjusted ≤ 0.05 CRP, C-Reactive protein. HFD, Healthy Food Diversity.

Supplementary table 3 Paired comparison of sample characteristics between time-points

Between Time-points (paired Wilcox test)

Meta-data	Groups	Comparison	Mean (SD)		P-value
			Group 1	Group 2	
Age	Control	T0 - T2	36.56 ± 9.88	36.89 ± 9.97	1.5E-01
		T0 - T3	38.14 ± 10.56	38.71 ± 10.8	7.2E-02
		T3 - T2	38.71 ± 10.8	38.57 ± 10.6	1.0E+00
	RA	T0 - T2	57 ± 7.21	57.17 ± 7.44	1.0E+00
		T0 - T3	57 ± 7.21	57.67 ± 7.31	7.2E-02
		T3 - T2	57.67 ± 7.31	57.17 ± 7.44	1.5E-01
	PSA	T0 - T2	48.75 ± 14.77	49.25 ± 14.75	3.5E-01
		T0 - T3	48 ± 18	48.67 ± 18.01	3.5E-01
		T3 - T2	48.67 ± 18.01	48.67 ± 18.01	NA
	AS	T0 - T2	47.67 ± 5.51	48 ± 6	1.0E+00
		T0 - T3	45 ± 4.24	45 ± 4.24	NA
		T3 - T2	45 ± 4.24	45 ± 4.24	NA
	IBD	T0 - T2	39.5 ± 11.96	40.25 ± 12.34	1.5E-01
		T0 - T3	35.67 ± 11.24	36.67 ± 11.24	1.5E-01
		T3 - T2	36.67 ± 11.24	36.33 ± 11.68	1.0E+00
BMI	Control	T0 - T2	26.49 ± 6.68	26.61 ± 6.72	2.7E-01
		T0 - T3	27.3 ± 7.47	27.86 ± 8.05	2.9E-01
		T3 - T2	27.86 ± 8.05	27.31 ± 7.57	**3.4E-02**
	RA	T0 - T2	27.53 ± 4.11	28.03 ± 3.92	6.9E-01
		T0 - T3	27.53 ± 4.11	27.63 ± 3.36	1.0E+00
		T3 - T2	27.63 ± 3.36	28.03 ± 3.92	2.9E-01
	PSA	T0 - T2	25.02 ± 7.17	25.12 ± 7.67	1.0E+00
		T0 - T3	22.23 ± 5.51	21.87 ± 5.16	2.5E-01
		T3 - T2	21.87 ± 5.16	21.97 ± 5.34	7.5E-01
	AS	T0 - T2	27.43 ± 3.33	28.1 ± 3.57	2.5E-01
		T0 - T3	26.9 ± 4.53	26.85 ± 4.45	1.0E+00
		T3 - T2	26.85 ± 4.45	27.25 ± 4.6	5.0E-01
	IBD	T0 - T2	27.83 ± 5.38	29.4 ± 3.21	2.5E-01
		T0 - T3	25.87 ± 4.52	28.4 ± 2.48	2.5E-01
		T3 - T2	28.4 ± 2.48	28.1 ± 2.31	1.0E+00
Height (m)	Control	T0 - T2	1.73 ± 0.13	1.73 ± 0.13	NA
		T0 - T3	1.71 ± 0.14	1.7 ± 0.13	1.7E-01
		T3 - T2	1.7 ± 0.13	1.71 ± 0.14	1.7E-01
	RA	T0 - T2	1.68 ± 0.10	1.68 ± 0.10	NA
		T0 - T3	1.68 ± 0.10	1.68 ± 0.10	NA
		T3 - T2	1.68 ± 0.10	1.68 ± 0.10	NA
	PSA	T0 - T2	1.73 ± 0.08	1.74 ± 0.08	1.0E+00
		T0 - T3	1.76 ± 0.06	1.77 ± 0.06	1.0E+00
		T3 - T2	1.77 ± 0.06	1.77 ± 0.06	NA

	AS	T0 - T2	1.70 ± 0.01	1.70 ± 0.01	NA
		T0 - T3	1.76 ± 0.09	1.76 ± 0.09	NA
		T3 - T2	1.76 ± 0.09	1.76 ± 0.09	NA
	IBD	T0 - T2	1.64 ± 0.10	1.64 ± 0.10	NA
		T0 - T3	1.65 ± 0.11	1.65 ± 0.11	NA
		T3 - T2	1.65 ± 0.11	1.65 ± 0.11	NA
HFD	Control	T0 - T2	0.51 ± 0.04	0.49 ± 0.04	6.5E-01
		T0 - T3	0.5 ± 0.05	0.52 ± 0.03	9.4E-01
		T3 - T2	0.52 ± 0.03	0.51 ± 0.03	5.8E-01
	RA	T0 - T2	0.31 ± 0.04	0.33 ± 0.07	3.1E-01
		T0 - T3	0.31 ± 0.04	0.37 ± 0.08	**3.1E-02**
		T3 - T2	0.37 ± 0.08	0.33 ± 0.07	**3.1E-02**
	PSA	T0 - T2	0.42 ± 0.06	0.43 ± 0.07	1.0E+00
		T0 - T3	0.39 ± 0.04	0.41 ± 0.07	1.0E+00
		T3 - T2	0.41 ± 0.07	0.42 ± 0.09	1.0E+00
	AS	T0 - T2	0.29 ± 0.01	0.31 ± 0.05	1.0E+00
		T0 - T3	0.3 ± 0	0.39 ± 0.02	5.0E-01
		T3 - T2	0.39 ± 0.02	0.33 ± 0.05	1.0E+00
	IBD	T0 - T2	0.29 ± 0.13	0.31 ± 0.13	1.8E-01
		T0 - T3	0.29 ± 0.16	0.35 ± 0.13	3.7E-01
		T3 - T2	0.35 ± 0.13	0.3 ± 0.16	3.7E-01
Weight (Kg)	Control	T0 - T2	79 ± 19.43	78.46 ± 17.32	6.8E-01
		T0 - T3	80.54 ± 22.05	80.46 ± 21.11	9.3E-01
		T3 - T2	80.46 ± 21.11	79.41 ± 19.74	9.4E-01
	RA	T0 - T2	78.73 ± 19.32	80.07 ± 18.93	4.0E-01
		T0 - T3	78.73 ± 19.32	78.82 ± 17.45	1.0E+00
		T3 - T2	78.82 ± 17.45	80.07 ± 18.93	1.6E-01
	PSA	T0 - T2	74.5 ± 18.16	75.08 ± 19.35	8.8E-01
		T0 - T3	69.43 ± 18.46	67.8 ± 17.66	2.5E-01
		T3 - T2	67.8 ± 17.66	69.1 ± 18.63	2.5E-01
	AS	T0 - T2	79.5 ± 17.51	80.73 ± 18.6	2.5E-01
		T0 - T3	84.15 ± 21.99	84.45 ± 22.7	1.0E+00
		T3 - T2	84.45 ± 22.7	85.8 ± 23.19	5.0E-01
	IBD	T0 - T2	73.53 ± 8.3	78.35 ± 5.36	2.5E-01
		T0 - T3	69.63 ± 3.51	77.13 ± 3.54	2.5E-01
		T3 - T2	77.13 ± 3.54	76.4 ± 4.5	1.0E+00
Total Physical activity	Control	T0 - T2	5536.78 ± 5173.11	5327.17 ± 4011.62	9.1E-01
		T0 - T3	6196.14 ± 5774.07	6169.83 ± 6374.05	3.1E-01
		T3 - T2	6169.83 ± 6374.05	5400.64 ± 4471.76	5.6E-01
	RA	T0 - T2	2407.5 ± 1898.96	4831.08 ± 4818.16	5.6E-01
		T0 - T3	2407.5 ± 1898.96	4305.58 ± 5789.11	4.4E-01

		T3 - T2	4305.58 ± 5789.11	4831.08 ± 4818.16	1.0E+00
	PSA	T0 - T2	12645.38 ± 19685.45	3759 ± 3123.53	2.5E-01
		T0 - T3	16394 ± 22292.66	7531 ± 1975.51	1.0E+00
		T3 - T2	7531 ± 1975.51	5490 ± 1238.85	1.0E+00
	AS	T0 - T2	6029 ± 3816.71	16424 ± 14126.85	5.0E-01
		T0 - T3	7744.5 ± 3387.75	28434 ± 17768.18	5.0E-01
		T3 - T2	28434 ± 17768.18	24576 ± 636.4	1.0E+00
	IBD	T0 - T2	372.67 ± 189.13	3270 ± 4951.75	5.0E-01
		T0 - T3	372.67 ± 189.13	1006.67 ± 1036.38	7.5E-01
		T3 - T2	1006.67 ± 1036.38	4100 ± 5713.7	2.5E-01
Morisky_score	RA	T0 - T2	5.88 ± 1.01	7.17 ± 1.17	5.9E-02
		T0 - T3	5.88 ± 1.01	8 ± 0	**3.5E-02**
		T3 - T2	8 ± 0	7.17 ± 1.17	1.7E-01
	PSA	T0 - T2	6.62 ± 1.16	6.44 ± 1.94	1.0E+00
		T0 - T3	7.17 ± 0.52	7 ± 1.32	1.0E+00
		T3 - T2	7 ± 1.32	7.25 ± 1.3	3.7E-01
	AS	T0 - T2	6.92 ± 1.13	8 ± 0	3.7E-01
		T0 - T3	6.88 ± 1.59	7.5 ± 0.71	1.0E+00
		T3 - T2	7.5 ± 0.71	8 ± 0	1.0E+00
	IBD	T0 - T2	5.38 ± 2.29	7.19 ± 0.55	1.8E-01
		T0 - T3	6.33 ± 1.53	7.67 ± 0.58	3.5E-01
		T3 - T2	7.67 ± 0.58	7.25 ± 0.66	1.0E+00
White blood cells	RA	T0 - T2	9.86 ± 2.4	7.4 ± 2.82	6.3E-02
		T0 - T3	9.86 ± 2.4	7.9 ± 2.99	6.3E-02
		T3 - T2	7.9 ± 2.99	7.4 ± 2.82	1.4E-01
	PSA	T0 - T2	6.25 ± 0.85	6 ± 1.93	8.8E-01
		T0 - T3	5.93 ± 0.7	6.77 ± 1.9	7.5E-01
		T3 - T2	6.77 ± 1.9	5.07 ± 0.61	2.5E-01
	AS	T0 - T2	7.63 ± 2.1	5.63 ± 0.84	2.5E-01
		T0 - T3	8.45 ± 2.19	5.6 ± 0.99	5.0E-01
		T3 - T2	5.6 ± 0.99	5.9 ± 0.99	3.5E-01
	IBD	T0 - T2	8.72 ± 2.19	7.2 ± 2.82	3.8E-01
		T0 - T3	9.6 ± 1.61	10.17 ± 4.05	7.5E-01
		T3 - T2	10.17 ± 4.05	7.23 ± 3.46	2.5E-01
Neutrophils	RA	T0 - T2	7 ± 2.07	4.35 ± 2.59	6.3E-02
		T0 - T3	7 ± 2.07	4.85 ± 2.86	6.3E-02
		T3 - T2	4.85 ± 2.86	4.35 ± 2.59	5.9E-02
	PSA	T0 - T2	3.87 ± 0.5	3.21 ± 1.31	3.8E-01
		T0 - T3	3.69 ± 0.44	4.29 ± 2.44	7.5E-01
		T3 - T2	4.29 ± 2.44	2.62 ± 0.65	2.5E-01
	AS	T0 - T2	5.27 ± 1.99	2.97 ± 0.55	2.5E-01
		T0 - T3	6.05 ± 2.05	2.9 ± 0.99	5.0E-01
		T3 - T2	2.9 ± 0.99	3.15 ± 0.64	1.0E+00

	IBD	T0 - T2	5.81 ± 1.63	4.3 ± 1.97	2.5E-01
		T0 - T3	6.48 ± 1.14	6.99 ± 3.78	7.5E-01
		T3 - T2	6.99 ± 3.78	4.44 ± 2.39	2.5E-01
Lymphocytes	RA	T0 - T2	1.78 ± 0.58	3.9 ± 4.52	6.3E-02
		T0 - T3	1.78 ± 0.58	2.05 ± 0.54	5.9E-01
		T3 - T2	2.05 ± 0.54	3.9 ± 4.52	8.3E-01
	PSA	T0 - T2	1.69 ± 0.39	1.98 ± 0.66	2.5E-01
		T0 - T3	1.55 ± 0.33	1.66 ± 0.39	1.0E+00
		T3 - T2	1.66 ± 0.39	1.71 ± 0.46	1.0E+00
	AS	T0 - T2	1.6 ± 0.1	1.93 ± 0.23	2.5E-01
		T0 - T3	1.6 ± 0.14	1.95 ± 0.07	5.0E-01
		T3 - T2	1.95 ± 0.07	2 ± 0.28	1.0E+00
	IBD	T0 - T2	1.84 ± 0.6	1.98 ± 0.9	6.3E-01
		T0 - T3	1.89 ± 0.72	1.79 ± 0.45	1.0E+00
		T3 - T2	1.79 ± 0.45	1.83 ± 1.04	1.0E+00
Monocytes	RA	T0 - T2	0.74 ± 0.11	0.5 ± 0.15	1.0E-01
		T0 - T3	0.74 ± 0.11	0.48 ± 0.18	1.0E-01
		T3 - T2	0.48 ± 0.18	0.5 ± 0.15	7.9E-01
	PSA	T0 - T2	0.39 ± 0.02	0.44 ± 0.12	4.2E-01
		T0 - T3	0.39 ± 0.02	0.49 ± 0.05	2.5E-01
		T3 - T2	0.49 ± 0.05	0.42 ± 0.14	3.7E-01
	AS	T0 - T2	0.5 ± 0.2	0.33 ± 0.06	3.7E-01
		T0 - T3	0.5 ± 0.28	0.4 ± 0.14	1.0E+00
		T3 - T2	0.4 ± 0.14	0.35 ± 0.07	1.0E+00
	IBD	T0 - T2	0.59 ± 0.13	0.43 ± 0.09	1.8E-01
		T0 - T3	0.63 ± 0.11	0.59 ± 0.11	1.0E+00
		T3 - T2	0.59 ± 0.11	0.44 ± 0.1	2.5E-01
Eosinophils	RA	T0 - T2	0.29 ± 0.29	16.63 ± 39.86	4.2E-01
		T0 - T3	0.29 ± 0.29	0.37 ± 0.22	3.7E-01
		T3 - T2	0.37 ± 0.22	16.63 ± 39.86	3.6E-01
	PSA	T0 - T2	0.17 ± 0.08	0.16 ± 0.05	7.9E-01
		T0 - T3	0.19 ± 0.08	0.17 ± 0.11	1.0E+00
		T3 - T2	0.17 ± 0.11	0.18 ± 0.03	1.0E+00
	AS	T0 - T2	0.13 ± 0.06	0.19 ± 0.08	2.5E-01
		T0 - T3	0.15 ± 0.07	0.18 ± 0.13	1.0E+00
		T3 - T2	0.18 ± 0.13	0.2 ± 0.11	5.0E-01
	IBD	T0 - T2	0.42 ± 0.47	0.37 ± 0.18	8.8E-01
		T0 - T3	0.52 ± 0.51	0.64 ± 0.4	3.7E-01
		T3 - T2	0.64 ± 0.4	0.39 ± 0.21	3.7E-01
Basophils	RA	T0 - T2	0.04 ± 0.05	5.52 ± 13.46	1.0E+00
		T0 - T3	0.04 ± 0.05	0.03 ± 0.04	1.0E+00

		T3 - T2	0.03 ± 0.04	5.52 ± 13.46	1.0E+00
	PSA	T0 - T2	0.03 ± 0.03	0.02 ± 0.02	7.9E-01
		T0 - T3	0.04 ± 0.03	0.05 ± 0.05	1.0E+00
		T3 - T2	0.05 ± 0.05	0.02 ± 0.02	1.0E+00
	AS	T0 - T2	0.05 ± 0.05	0.06 ± 0.01	7.5E-01
		T0 - T3	0.05 ± 0.07	0.06 ± 0.02	1.0E+00
		T3 - T2	0.06 ± 0.02	0.06 ± 0.01	1.0E+00
	IBD	T0 - T2	0.04 ± 0.04	0.04 ± 0.05	3.5E-01
		T0 - T3	0.05 ± 0.04	0.09 ± 0.01	3.7E-01
		T3 - T2	0.09 ± 0.01	0.05 ± 0.05	3.7E-01
CRP	RA	T0 - T2	18.4 ± 14.61	6.18 ± 9.25	**3.1E-02**
		T0 - T3	18.4 ± 14.61	11.84 ± 15.86	3.1E-01
		T3 - T2	11.84 ± 15.86	6.18 ± 9.25	2.5E-01
	PSA	T0 - T2	2.01 ± 1.39	1.03 ± 1.7	1.0E+00
		T0 - T3	1.68 ± 1.5	2.02 ± 1.98	1.0E+00
		T3 - T2	2.02 ± 1.98	0.05 ± 0	1.0E+00
	AS	T0 - T2	4.67 ± 2.08	0.7 ± 1.13	2.5E-01
		T0 - T3	3.5 ± 0.71	0.05 ± 0	5.0E-01
		T3 - T2	0.05 ± 0	0.05 ± 0	NaN
	IBD	T0 - T2	13.7 ± 17.51	2.73 ± 1.74	1.8E-01
		T0 - T3	16.57 ± 20.26	11.9 ± 19.14	7.5E-01
		T3 - T2	11.9 ± 19.14	2.63 ± 2.12	1.0E+00
Disease activity	RA	T0 - T2	5.3 ± 0.25	4.24 ± 0.86	1.3E-01
		T0 - T3	5.3 ± 0.25	4.08 ± 1.08	1.3E-01
		T3 - T2	4.08 ± 1.08	4.24 ± 0.86	6.3E-01
	PSA	T0 - T2	18.82 ± 8.75	19.75 ± 16.44	1.0E+00
		T0 - T3	15.83 ± 7.82	9 ± 4.36	2.5E-01
		T3 - T2	9 ± 4.36	13.33 ± 12.58	5.9E-01
	AS	T0 - T2			
		T0 - T3	Not performed		
		T3 - T2			
	IBD	T0 - T2	5 ± 4.32	2.67 ± 2.31	7.5E-01
		T0 - T3	6.33 ± 4.16	0.5 ± 0.71	5.0E-01
		T3 - T2	0.5 ± 0.71	2 ± 2.83	1.0E+00

Paired Wilcoxon test between the time-point values of different meta-data values in different disease and control groups to identify variation in time-point within the groups. Significance is defined by nominal p-values ≤ 0.05.

Supplementary table 4 Significance of comparison of sample characteristics of Time-point disease samples with controls and Established samples

A) Result of comparison of time-point samples of each disease group against controls

	Age	BMI	Height	Weight	Total physical activity	HFD
Control_T0 - AS_T0	1.9E-01	8.2E-01	**8.7E-03**	9.2E-01	1.5E-01	**2.9E-02**
Control_T0 - AS_T2	1.9E-01	8.2E-01	**1.3E-02**	9.2E-01	4.2E-01	**6.2E-03**
Control_T0 - AS_T3	2.3E-01	8.2E-01	**2.9E-02**	9.2E-01	7.8E-02	**2.0E-02**
Control_T0 - PSA_T0	5.2E-02	7.5E-01	**6.2E-04**	8.1E-01	8.0E-01	**2.3E-03**
Control_T0 - PSA_T2	1.8E-01	1.0E+00	**4.3E-03**	1.0E+00	1.0E+00	6.4E-02
Control_T0 - PSA_T3	2.2E-01	8.4E-01	**8.5E-03**	8.1E-01	1.9E-01	6.4E-02
Control_T0 - RA_T0	**1.8E-03**	3.4E-01	**1.5E-04**	8.5E-01	9.3E-01	**4.4E-05**
Control_T0 - RA_T2	**2.2E-03**	3.0E-01	**4.6E-04**	8.5E-01	9.3E-01	**4.4E-05**
Control_T0 - RA_T3	**2.2E-03**	3.0E-01	**4.6E-04**	8.5E-01	9.3E-01	**3.4E-03**
Control_T0 - IBD_T0	9.0E-01	4.8E-01	**4.3E-03**	1.0E+00	**6.2E-03**	**1.7E-02**
Control_T0 - IBD_T2	9.0E-01	4.4E-01	**4.3E-03**	1.0E+00	1.2E-01	**1.7E-02**
Control_T0 - IBD_T3	9.1E-01	4.8E-01	**8.5E-03**	1.0E+00	7.1E-02	**4.7E-02**

B) Comparison between Time-point disease samples and Established samples

	Ankylosing Spondylitis			**Psoriatic arthritis**			**Rheumatoid arthritis**		
	AS vs AS_T0	**AS vs AS_T2**	**AS vs AS_T3**	**PSA vs PSA_T0**	**PSA vs PSA_T2**	**PSA vs PSA_T3**	**RA vs RA_T0**	**RA vs RA_T2**	**RA vs RA_T3**
Age	7.1E-01	1.0E+00	7.1E-01	1.0E+00	1.0E+00	1.0E+00	5.3E-01	5.3E-01	5.3E-01
BMI	5.6E-01	6.1E-01	5.8E-01	7.4E-01	3.5E-01	1.3E-01	9.1E-01	9.1E-01	9.1E-01
Height (m)	8.3E-01	8.3E-01	8.3E-01	4.1E-01	2.8E-01	2.8E-01	3.8E-01	3.8E-01	3.8E-01
Weight (KG)	4.3E-01	7.0E-01	9.1E-01	9.8E-01	5.2E-01	5.2E-01	8.5E-01	8.5E-01	8.5E-01
Total physical activity	1.0E+00	1.0E+00	9.8E-01	6.2E-01	7.2E-01	2.4E-01	4.9E-01	4.4E-01	4.4E-01
Morisky score	6.1E-01	1.3E-01	4.0E-01	9.3E-01	9.3E-01	9.3E-01	8.7E-02	1.0E+00	1.0E-01
White blood cells	9.3E-01	9.3E-01	9.3E-01	9.2E-01	9.2E-01	9.2E-01	**2.9E-02**	7.6E-01	4.4E-01
Neutrophils	6.8E-01	1.0E+00	1.0E+00	4.8E-01	4.8E-01	8.8E-01	**8.0E-03**	9.7E-01	9.2E-01
Lymphocytes	2.0E-01	9.1E-01	9.1E-01	7.7E-01	7.7E-01	7.7E-01	1.9E-01	1.9E-01	5.1E-01
Monocytes	8.6E-01	8.6E-01	8.6E-01	4.2E-01	5.4E-01	4.2E-01	**1.8E-02**	3.6E-01	4.5E-01
Eosinophils	1.9E-01	7.5E-01	7.5E-01	1.0E+00	1.0E+00	1.0E+00	7.8E-01	5.6E-01	4.5E-01
Basophils	7.7E-01	2.9E-01	2.9E-01	7.8E-01	3.1E-01	1.0E+00	1.0E+00	4.8E-01	4.8E-01
CRP	6.6E-01	4.3E-01	4.3E-01	6.0E-01	6.0E-01	9.1E-01	**2.1E-02**	9.2E-01	3.6E-01
HFD	2.4E-01	2.4E-01	7.6E-01	5.7E-01	5.7E-01	5.7E-01	9.7E-01	9.7E-01	9.7E-01

Supplementary table 4A) represents P-adjusted values from the pairwise comparison between the meta-data values of the disease at particular time-point against Control samples at T0. Supplementary table 4B) represents comparison of various time-point diseased samples against the established treated group for the same disease. Significance is defined by P-adjusted ≤ 0.05

Supplementary table 5 Within sample taxonomic and functional indices

Groups	Gene count	Unique_species	Shannon_diversity
control_T0	200673.8 ± 25193.01	73.31 ± 9.25	2.98 ± 0.23
control_T2	195856.4 ± 41917.72	72.11 ± 12.52	2.84 ± 0.48
control_T3	191205.9 ± 48288.58	70.29 ± 16.95	2.82 ± 0.35
RA	204749.8 ± 27963.77	82.81 ± 10.23	3.00 ± 0.23
RA_T0	217989.9 ± 33998.53	87.78 ± 21.28	3.00 ± 0.21
RA_T2	203455.3 ± 25000.44	81.67 ± 7.89	2.74 ± 0.26
RA_T3	206872 ± 34865.97	87.33 ± 12.69	3.06 ± 0.35
AS	183013.6 ± 48421.07	69.80 ± 17.94	2.64 ± 0.60
AS_T0	215356.5 ± 18421.81	86.00 ± 7.87	3.07 ± 0.12
AS_T2	212473 ± 28490.29	80.50 ± 8.54	3.02 ± 0.08
AS_T3	199550 ± 4690.95	87.50 ± 13.44	3.02 ± 0.12
PSA	206789.9 ± 26726.11	75.78 ± 11.20	3.07 ± 0.29
PSA_T0	199940.7 ± 15200.11	77.00 ± 6.32	2.90 ± 0.33
PSA_T2	194479.8 ± 10286.26	73.75 ± 4.79	2.87 ± 0.02
PSA_T3	176107 ± 50960.77	64.33 ± 13.32	2.91 ± 0.24
IBD_T0	115160.8 ± 31350.50	56.75 ± 13.18	2.16 ± 0.28
IBD_T2	126724 ± 48964.62	57.75 ± 28.93	2.32 ± 0.31
IBD_T3	149461.7 ± 62444.76	63.33 ± 30.24	2.52 ± 0.43

Mean and standard deviation of different within-sample indices observed across different groups including time-points and established samples.

Supplementary table 6 Comparison of within sample diversity within time-points, against controls and established groups

A) Significance of paired comparison between time-point of within sample indices

	P-values		
	Unique species	**Shannon**	**gene count**
control_T2 vs T0	1.0E+00	8.2E-01	9.1E-01
control_T3 vs T0	8.3E-01	1.1E-01	5.8E-01
control_T3 vs T2	7.3E-01	6.9E-01	6.9E-01
RA_T2 vs T0	4.2E-01	**6.3E-02**	3.1E-01
RA_T3 vs T0	4.4E-01	8.4E-01	8.4E-01
RA_T3 vs T2	5.6E-01	**9.4E-02**	6.9E-01
PSA_T2 vs T0	4.2E-01	2.5E-01	1.0E+00
PSA_T3 vs T0	3.7E-01	5.0E-01	7.5E-01
PSA_T3 vs T2	7.5E-01	7.5E-01	1.0E+00
AS_T2 vs T0	1.3E-01	6.3E-01	8.8E-01
AS_T3 vs T0	1.0E+00	1.0E+00	1.0E+00
AS_T3 vs T2	5.0E-01	1.0E+00	5.0E-01
IBD_T2 vs T0	8.8E-01	6.3E-01	1.0E+00
IBD_T3 vs T0	5.0E-01	2.5E-01	5.0E-01
IBD_T3 vs T2	1.0E+00	7.5E-01	5.0E-01

P-adjusted values is based on all pairwise comparison between time-point for a particular group. Within-sample diversity was not significant within different time-point of each disease and healthy groups. Significance is defined by P-adjusted values (≤ 0.05) of all pairwise comparisons with each group. Trends of association is defined as P-adjusted values ≤ 0.1

B) Significance of within sample diversity between disease time-point against control T0 samples

	Unique species	Shannon	gene count
Control_T0 vs AS_T0	**1.8E-02**	6.8E-01	4.4E-01
Control_T0 vs AS_T2	2.2E-01	9.6E-01	4.4E-01
Control_T0 vs AS_T3	1.6E-01	1.0E+00	8.4E-01
Control_T0 vs IBD_T0	**1.6E-02**	**4.1E-04**	**8.3E-04**
Control_T0 vs IBD_T2	4.2E-01	**2.9E-03**	**2.9E-03**
Control_T0 vs IBD_T3	7.8E-01	**6.4E-02**	1.1E-01
Control_T0 vs PSA_T0	4.4E-01	7.7E-01	7.7E-01
Control_T0 vs PSA_T2	7.4E-01	1.5E-01	6.2E-01
Control_T0 vs PSA_T3	2.2E-01	5.6E-01	5.6E-01
Control_T0 vs RA_T0	**7.9E-02**	8.0E-01	2.3E-01
Control_T0 vs RA_T2	**3.2E-02**	1.1E-01	8.6E-01
Control_T0 vs RA_T3	**2.2E-02**	4.9E-01	5.9E-01
Control_T0 vs AS	4.3E-01	**8.7E-02**	4.8E-01
Control_T0 vs PSA	1.7E-01	2.2E-01	4.8E-01
Control_T0 vs RA	**1.6E-02**	8.7E-01	7.2E-01

Pairwise comparison of different time-point and established samples from diseased groups against controls. Significance is defined by P-adjusted values (≤ 0.05) between all pairwise comparisons of all time-points from a disease group against Control T0. Trends of association is defined as P-adjusted values ≤ 0.1

Supplementary table 7 Comparison of taxonomic β-diversity between various groups

		Spearman Distance	
		Observed estimate	P-value
	Between controls	0.05	9.1E-01
Established Groups & Controls	**AS**	0.06	1.6E-01
	PSA	0.04	2.5E-01
	RA	0.05	1.1E-01
Time-point Groups & Controls	**AS**	0.16	**4.3E-02**
	PSA	0.10	7.8E-01
	RA	0.10	4.6E-01
	IBD	0.38	**1.0E-03**
Timeseries, established & controls	**RA**	0.09	4.0E-01
	AS	0.15	**4.0E-02**
	PSA	0.08	8.0E-01
Between various diseases	**RA & IBD**	0.25	**1.0E-03**
	RA & AS	0.04	**3.0E-03**
	RA & PSA	0.03	**6.0E-03**
	PSA & IBD	0.29	**1.0E-03**
	PSA & AS	0.03	1.3E-01
	AS & IBD	0.29	**1.0E-03**
Independent dataset	**Zhang only**	0.05	**1.0E-03**
	Zhang & RA Biologics	0.12	**1.0E-03**

Significance based on Monte-Carlo permutations test from the BCA analysis of individually compared time-point groups against controls, established samples against controls, timeseries and established together against controls, comparison of the different disease samples amongst themselves and validating the significance in the public datasets. Significance is defined by nominal p-value ≤ 0.05

Supplementary table 8 Comparison of β-diversity between timepoints observed in timeseries groups

A) summary and analysis of variance (ANOVA) from the linear model of dissimilarity values as explained by group stratification adjusted for observed species

			Summary			ANOVA
			Estimate	Std. Error	R-squared	P-value
Spearman	T2-T0	Intercept	0.22	0.09	0.40	1.3E-01
		Observed species	0.00	0.00		
		control	0.00	0.04		
		IBD	0.09	0.05		
		PSA	-0.05	0.05		
		RA	0.02	0.05		
	T3-T0	Intercept	0.34	0.05	0.78	**4.1E-04**
		Observed species	0.00	0.00		
		control	-0.06	0.03		
		IBD	0.03	0.03		
		PSA	-0.09	0.03		
		RA	0.05	0.03		

The model investigated here is:

The distance between paired time-points samples ~ observed species + Groups

Significance as defined by nominal P-values (≤ 0.05) are highlighted in red and bold.

B) Pairwise comparison of Spearman dissimilarity based on significance as defined by ANOVA analysis on models of **Supplementary table 8A**.

Spearman	T2-T0				T3-T0			
	difference	Lower interval	Upper interval	P-adjusted	difference	Lower interval	Upper interval	P-adjusted
control-AS	-0.01	-0.13	0.11	1.0E+00	-0.04	-0.13	0.05	6.4E-01
IBD-AS	0.07	-0.07	0.22	5.8E-01	0.05	-0.05	0.16	4.9E-01
PSA-AS	-0.05	-0.20	0.09	8.2E-01	-0.07	-0.17	0.04	3.1E-01
RA-AS	0.02	-0.11	0.15	9.9E-01	0.05	-0.04	0.14	5.3E-01
IBD-control	0.08	-0.04	0.21	3.0E-01	0.09	0.02	0.17	**1.3E-02**
PSA-control	-0.04	-0.17	0.08	8.4E-01	-0.03	-0.10	0.05	8.4E-01
RA-control	0.03	-0.08	0.14	9.1E-01	0.09	0.02	0.15	**4.6E-03**
PSA-IBD	-0.13	-0.27	0.02	1.2E-01	-0.12	-0.21	-0.03	**7.4E-03**
RA-IBD	-0.05	-0.19	0.08	7.7E-01	-0.01	-0.09	0.07	1.0E+00
RA-PSA	0.07	-0.06	0.21	4.9E-01	0.11	0.03	0.19	**4.0E-03**

Significance as defined by adjusted P-values (≤ 0.05) are highlighted in red and bold.

Supplementary table 9 Cumulative comparison of functional β-diversity by including established and time-point samples of each disease type and controls

FUNCTIONAL	Spearman	
	Observed estimate	P-value
RA time-point, Establish & Control	0.10	2.1E-01
AS time-point, Establish & Control	0.13	2.2E-01
PSA time-point, Establish & Control	0.10	2.7E-01
IBD time-point & Control	0.37	1.0e-03

Supplementary table 10 Pairwise results of taxonomic differential abundance analysis

A) Significant pairwise comparisons of rheumatoid arthritis (RA) time-point, established samples and controls T0

RA T0 & Control T0	log2FoldChange	lfcSE	pvalue	padj
Bacteroides_salyersiae	-30.00	3.58	5.5E-17	3.4E-15
Dialister_invisus	-30.00	4.43	1.3E-11	3.8E-10
Slackia_piriformis	-26.34	4.40	2.2E-09	2.7E-08
Blautia_hydrogenotrophica	-25.71	4.43	6.4E-09	6.5E-08
Bacteroides_clarus	-23.73	4.43	8.4E-08	7.9E-07
Streptococcus_infantis	23.17	4.41	1.5E-07	1.3E-06
Lactobacillus_casei_paracasei	25.82	4.40	4.3E-09	4.7E-08
Actinomyces_viscosus	27.09	4.41	7.8E-10	1.1E-08
Veillonella_dispar	27.59	4.41	3.8E-10	5.8E-09
Streptococcus_gordonii	27.80	4.38	2.2E-10	3.9E-09
Rothia_dentocariosa	28.50	4.41	9.9E-11	2.0E-09
Collinsella_intestinalis	28.78	4.29	2.0E-11	5.0E-10
Acidaminococcus_intestini	29.95	4.41	1.1E-11	3.8E-10
Streptococcus_anginosus	30.00	2.99	1.1E-23	1.4E-21
RA T2 & Control T0				
Bacteroides_salyersiae	-30.00	4.13	3.5E-13	1.4E-11
Dialister_invisus	-30.00	5.10	3.9E-09	6.9E-08
Holdemania_unclassified	-28.65	3.16	1.3E-19	1.6E-17
Erysipelotrichaceae_bacterium_6_1_45	-26.86	3.81	1.7E-12	5.2E-11
Anaerotruncus_unclassified	-26.35	5.10	2.3E-07	2.4E-06
Slackia_piriformis	-24.75	5.07	1.0E-06	8.4E-06
Actinomyces_odontolyticus	-24.54	3.83	1.4E-10	3.5E-09
Anaerotruncus_colihominis	-23.79	4.52	1.4E-07	1.7E-06
Bacteroides_clarus	-23.25	5.10	5.1E-06	3.7E-05
Blautia_hydrogenotrophica	-23.23	5.10	5.2E-06	3.7E-05
Lactobacillus_casei_paracasei	20.65	5.04	4.3E-05	2.9E-04
Acidaminococcus_intestini	24.87	5.05	8.6E-07	7.5E-06
Streptococcus_gordonii	25.24	5.03	5.1E-07	4.8E-06
Actinomyces_viscosus	26.44	5.05	1.7E-07	1.9E-06
Streptococcus_infantis	26.92	5.05	1.0E-07	1.4E-06
Streptococcus_mutans	27.31	5.05	6.5E-08	9.9E-07
Streptococcus_anginosus	28.83	3.42	3.8E-17	2.3E-15
Collinsella_intestinalis	30.00	4.92	1.1E-09	2.3E-08
RA T3 & Control T0				
Bacteroides_salyersiae	-30.00	4.13	3.5E-13	1.4E-11
Holdemania_unclassified	-29.40	3.16	1.4E-20	1.8E-18
Lactococcus_lactis	-28.89	4.52	1.6E-10	3.8E-09
Actinomyces_odontolyticus	-25.51	3.83	2.6E-11	8.1E-10
Pseudoflavonifractor_capillosus	-24.82	4.77	2.0E-07	1.9E-06
Blautia_hydrogenotrophica	-24.09	5.10	2.3E-06	1.9E-05
Bacteroides_clarus	-23.98	5.10	2.5E-06	1.9E-05
Streptococcus_gordonii	23.36	5.03	3.4E-06	2.4E-05

Acidaminococcus_intestini	25.64	5.05	3.9E-07	3.4E-06
Streptococcus_infantis	26.87	5.05	1.1E-07	1.1E-06
Veillonella_dispar	27.21	5.05	7.3E-08	8.1E-07
Streptococcus_mutans	27.62	5.05	4.6E-08	5.6E-07
Rothia_dentocariosa	28.64	5.05	1.4E-08	2.0E-07
Collinsella_intestinalis	28.98	4.92	4.0E-09	8.1E-08
Actinomyces_viscosus	29.13	5.05	8.2E-09	1.3E-07
Streptococcus_anginosus	29.23	3.42	1.4E-17	8.5E-16
Lactobacillus_casei_paracasei	29.52	5.04	4.8E-09	8.4E-08
RA & Control T0				
Lachnospiraceae_bacterium_1_4_56FAA	7.36	2.50	3.2E-03	2.8E-02
Actinomyces_viscosus	23.43	3.75	4.1E-10	4.5E-09
Rothia_dentocariosa	26.12	3.75	3.2E-12	3.9E-11
Acidaminococcus_intestini	26.21	3.75	2.7E-12	3.7E-11
Streptococcus_infantis	26.59	3.75	1.3E-12	2.0E-11
Streptococcus_mutans	27.13	3.75	4.5E-13	7.9E-12
Veillonella_dispar	27.15	3.75	4.4E-13	7.9E-12
Streptococcus_anginosus	28.15	2.55	2.8E-28	3.4E-26
Streptococcus_gordonii	28.26	3.73	3.4E-14	8.4E-13
Collinsella_intestinalis	29.40	3.65	8.5E-16	3.7E-14
Alistipes_sp_AP11	30.00	3.75	1.2E-15	3.7E-14
Lactobacillus_casei_paracasei	30.00	3.74	1.0E-15	3.7E-14
RA T3 & RA T0	**log2FoldChange**	**lfcSE**	**pvalue**	**padj**
Actinomyces_odontolyticus	-29.78	4.20	1.3E-12	7.8E-11
Pseudoflavonifractor_capillosus	-29.12	5.24	2.7E-08	5.5E-07
Lactococcus_lactis	-28.89	4.96	5.6E-09	1.4E-07
Holdemania_unclassified	-24.92	3.46	6.0E-13	7.3E-11
Dialister_invisus	25.01	5.58	7.3E-06	1.3E-04
Slackia_piriformis	32.30	5.54	5.7E-09	1.4E-07
Streptococcus_mutans	36.71	5.58	4.6E-11	1.9E-09
RA T2 & RA T0				
Rothia_dentocariosa	-37.76	5.60	1.5E-11	6.2E-10
Actinomyces_odontolyticus	-28.81	4.20	6.6E-12	4.0E-10
Anaerotruncus_unclassified	-28.67	5.60	3.0E-07	5.3E-06
Veillonella_dispar	-28.30	5.60	4.3E-07	6.6E-06
Anaerotruncus_colihominis	-28.19	4.96	1.3E-08	2.7E-07
Erysipelotrichaceae_bacterium_6_1_45	-27.91	4.17	2.3E-11	6.9E-10
Holdemania_unclassified	-24.17	3.46	2.9E-12	3.5E-10
Streptococcus_mutans	36.40	5.58	6.7E-11	1.6E-09
RA T3 & RA T2				
Lactococcus_lactis	-28.99	5.42	8.7E-08	3.6E-06
Pseudoflavonifractor_capillosus	-28.64	5.73	5.7E-07	1.2E-05
Dialister_invisus	25.01	6.12	4.4E-05	6.0E-04
Erysipelotrichaceae_bacterium_6_1_45	25.66	4.56	1.8E-08	1.1E-06
Anaerotruncus_colihominis	27.31	5.42	4.7E-07	1.2E-05
Veillonella_dispar	27.92	6.12	5.1E-06	7.7E-05
Anaerotruncus_unclassified	29.28	6.12	1.7E-06	3.0E-05
Slackia_piriformis	30.72	6.08	4.5E-07	1.2E-05
Rothia_dentocariosa	37.90	6.12	5.9E-10	7.2E-08
RA T0 & RA	**log2FoldChange**	**lfcSE**	**pvalue**	**padj**
Streptococcus_mutans	-36.22	4.43	2.9E-16	1.8E-14
Alistipes_sp_AP11	-33.96	4.43	1.8E-14	7.1E-13
Slackia_piriformis	-32.83	4.40	8.9E-14	2.7E-12
Bacteroides_salyersiae	-29.34	3.58	2.6E-16	1.8E-14
Blautia_hydrogenotrophica	-28.73	4.43	8.7E-11	2.1E-09

Dialister_invisus	-26.50	4.43	2.2E-09	4.5E-08
Bacteroides_clarus	-24.81	4.43	2.1E-08	3.7E-07
RA T2 & RA				
Rothia_dentocariosa	-35.39	5.10	3.8E-12	9.3E-11
Alistipes_sp_AP11	-33.47	5.10	5.1E-11	1.0E-09
Slackia_piriformis	-31.24	5.07	7.0E-10	1.1E-08
Bacteroides_salyersiae	-29.34	4.13	1.1E-12	3.5E-11
Erysipelotrichaceae_bacterium_6_1_45	-28.13	3.81	1.5E-13	8.9E-12
Anaerotruncus_colihominis	-28.05	4.52	5.5E-10	9.6E-09
Actinomyces_odontolyticus	-28.01	3.83	2.5E-13	1.0E-11
Veillonella_dispar	-27.86	5.10	4.6E-08	6.2E-07
Holdemania_unclassified	-27.06	3.16	1.2E-17	1.4E-15
Dialister_invisus	-26.50	5.10	2.0E-07	2.4E-06
Anaerotruncus_unclassified	-26.33	5.10	2.4E-07	2.6E-06
Blautia_hydrogenotrophica	-26.25	5.10	2.6E-07	2.6E-06
Bacteroides_clarus	-24.32	5.10	1.8E-06	1.7E-05
Lachnospiraceae_bacterium_5_1_63FAA	3.65	1.16	1.7E-03	1.5E-02
RA T3 & RA				
Alistipes_sp_AP11	-34.40	5.10	1.5E-11	4.5E-10
Bacteroides_salyersiae	-29.34	4.13	1.1E-12	4.7E-11
Lactococcus_lactis	-29.21	4.52	9.9E-11	2.4E-09
Actinomyces_odontolyticus	-28.98	3.83	3.7E-14	2.3E-12
Holdemania_unclassified	-27.81	3.16	1.5E-18	1.8E-16
Blautia_hydrogenotrophica	-27.11	5.10	1.0E-07	1.8E-06
Pseudoflavonifractor_capillosus	-26.61	4.77	2.5E-08	5.0E-07
Bacteroides_clarus	-25.06	5.10	8.8E-07	1.3E-05

Significance is defined by P-adjusted ≤ 0.05. Red colour indicates species with reduced abundance while green represents species with increased abundance in that pairwise comparison.

B) Significant pairwise comparisons of ankylosing spondylitis (AS) time-point, established samples and controls T0

AS T0 & Control T0	**log2FoldChange**	**lfcSE**	**pvalue**	**padj**
Desulfovibrio_piger	-36.28	4.65	6.1E-15	3.6E-13
Bacteroides_finegoldii	-34.46	4.91	2.2E-12	6.7E-11
Bacteroides_massiliensis	-30.94	4.03	1.7E-14	6.6E-13
Desulfovibrio_desulfuricans	-30.88	4.91	3.2E-10	5.4E-09
Acidaminococcus_unclassified	-30.78	4.91	3.6E-10	5.4E-09
Coprobacter_fastidiosus	-30.52	4.91	5.1E-10	6.8E-09
Mitsuokella_unclassified	-30.41	4.91	5.9E-10	7.0E-09
Bacteroides_salyersiae	-30.32	4.91	6.6E-10	7.2E-09
Bacteroides_coprocola	-30.02	4.91	9.7E-10	9.6E-09
Mitsuokella_multacida	-29.53	4.91	1.8E-09	1.6E-08
Eubacterium_biforme	-29.40	4.61	1.9E-10	3.7E-09
Bacteroides_clarus	-28.69	4.91	5.1E-09	4.3E-08
Olsenella_unclassified	-28.47	4.91	6.7E-09	4.7E-08
Bifidobacterium_bifidum	-28.34	4.88	6.4E-09	4.7E-08
Bacteroides_faecis	-28.21	4.91	9.2E-09	5.8E-08
Lactococcus_lactis	-28.14	4.86	7.2E-09	4.7E-08
Megasphaera_unclassified	-26.76	4.91	5.1E-08	3.0E-07
Bacteroides_stercoris	-10.53	3.78	5.3E-03	3.0E-02
Clostridium_hathewayi	8.34	3.01	5.6E-03	3.1E-02
Anaerotruncus_colihominis	28.00	4.83	6.9E-09	4.7E-08
Clostridium_nexile	32.73	4.83	1.3E-11	3.0E-10
Collinsella_intestinalis	43.59	4.30	3.7E-24	4.4E-22
AS T2 & Control T0				

Desulfovibrio_piger	**-36.04**	4.65	9.2E-15	5.5E-13
Bacteroides_finegoldii	**-34.53**	4.91	2.0E-12	6.0E-11
Bacteroides_massiliensis	**-30.94**	4.03	1.7E-14	6.6E-13
Desulfovibrio_desulfuricans	**-30.87**	4.91	3.2E-10	7.1E-09
Odoribacter_unclassified	**-30.79**	4.91	3.6E-10	7.1E-09
Acidaminococcus_unclassified	**-30.59**	4.91	4.6E-10	7.7E-09
Bacteroides_salyersiae	**-30.51**	4.91	5.2E-10	7.7E-09
Mitsuokella_unclassified	**-30.26**	4.91	7.1E-10	9.4E-09
Bacteroides_coprocola	**-30.02**	4.91	9.7E-10	1.2E-08
Mitsuokella_multacida	**-29.70**	4.91	1.5E-09	1.6E-08
Olsenella_unclassified	**-28.42**	4.91	7.1E-09	5.6E-08
Bifidobacterium_bifidum	**-28.34**	4.88	6.4E-09	5.6E-08
Bacteroides_clarus	**-28.33**	4.91	7.9E-09	5.9E-08
Lactococcus_lactis	**-28.28**	4.86	6.0E-09	5.6E-08
Bacteroides_faecis	**-28.16**	4.91	9.8E-09	6.8E-08
Ruminococcaceae_bacterium_D16	**-27.56**	4.91	2.0E-08	1.2E-07
Megasphaera_unclassified	**-26.76**	4.91	5.1E-08	3.0E-07
Bifidobacterium_animalis	**-26.51**	4.72	2.0E-08	1.2E-07
Clostridium_hathewayi	**10.35**	3.01	5.9E-04	3.2E-03
Clostridium_nexile	**23.50**	4.83	1.2E-06	6.6E-06
Anaerotruncus_colihominis	**27.98**	4.83	7.0E-09	5.6E-08
Collinsella_intestinalis	**38.44**	4.30	3.9E-19	4.6E-17
AS T3 & Control T0				
Bacteroides_stercoris	**-32.61**	5.25	5.4E-10	1.1E-08
Bacteroides_salyersiae	**-32.39**	6.58	8.7E-07	6.9E-06
Desulfovibrio_piger	**-31.69**	6.24	3.8E-07	3.5E-06
Bacteroides_finegoldii	**-31.14**	6.58	2.3E-06	1.6E-05
Bacteroides_massiliensis	**-30.94**	5.41	1.1E-08	1.9E-07
Odoribacter_splanchnicus	**-30.69**	6.58	3.2E-06	2.1E-05
Mitsuokella_multacida	**-30.58**	6.58	3.4E-06	2.1E-05
Alistipes_indistinctus	**-30.48**	4.34	2.2E-12	6.5E-11
Bacteroides_coprocola	**-30.02**	6.58	5.1E-06	3.0E-05
Odoribacter_unclassified	**-29.90**	6.58	5.6E-06	3.2E-05
Paraprevotella_clara	**-29.88**	5.78	2.3E-07	2.3E-06
Paraprevotella_unclassified	**-29.82**	5.32	2.0E-08	2.8E-07
Bacteroides_cellulosilyticus	**-29.75**	5.31	2.1E-08	2.8E-07
Eubacterium_ventriosum	**-29.69**	3.07	3.6E-22	4.3E-20
Ruminococcus_bromii	**-29.55**	4.26	3.9E-12	9.2E-11
Eubacterium_biforme	**-29.40**	6.19	2.1E-06	1.5E-05
Lachnospiraceae_bacterium_2_1_58FAA	**-29.16**	5.46	9.1E-08	1.1E-06
Alistipes_unclassified	**-28.92**	6.58	1.1E-05	5.8E-05
Bilophila_wadsworthia	**-28.88**	3.55	3.8E-16	2.2E-14
Bifidobacterium_bifidum	**-28.34**	6.54	1.5E-05	7.4E-05
Burkholderiales_bacterium_1_1_47	**-28.13**	6.56	1.8E-05	8.7E-05
Acidaminococcus_unclassified	**-27.91**	6.58	2.2E-05	1.0E-04
Clostridium_citroniae	**-27.80**	6.58	2.4E-05	1.1E-04
Bacteroides_faecis	**-27.66**	6.58	2.7E-05	1.1E-04
Eggerthella_lenta	**-27.33**	5.19	1.4E-07	1.5E-06
Coprobacter_fastidiosus	**-27.31**	6.58	3.4E-05	1.3E-04
Desulfovibrio_desulfuricans	**-27.17**	6.58	3.7E-05	1.4E-04
Mitsuokella_unclassified	**-26.76**	6.58	4.8E-05	1.7E-04
Megasphaera_unclassified	**-26.76**	6.58	4.8E-05	1.7E-04
Bifidobacterium_animalis	**-26.51**	6.33	2.8E-05	1.2E-04
Ruminococcaceae_bacterium_D16	**-26.09**	6.58	7.4E-05	2.6E-04
Flavonifractor_plautii	**-25.73**	5.19	7.0E-07	5.9E-06
Bacteroides_clarus	**-25.14**	6.58	1.3E-04	4.6E-04
Olsenella_unclassified	**-24.90**	6.59	1.6E-04	5.2E-04
Pseudoflavonifractor_capillosus	**-24.86**	6.58	1.6E-04	5.2E-04
Parabacteroides_distasonis	**-6.88**	2.62	8.7E-03	2.7E-02

Oscillibacter_unclassified	-5.29	1.99	7.8E-03	2.4E-02
Clostridium_nexile	28.34	6.45	1.1E-05	5.8E-05
Collinsella_intestinalis	42.22	5.73	1.8E-13	7.0E-12
AS & Control T0				
Desulfovibrio_piger	-30.00	3.35	3.7E-19	4.5E-17
Mitsuokella_multacida	-27.09	3.54	2.1E-14	4.7E-13
Odoribacter_unclassified	-27.03	3.54	2.4E-14	4.7E-13
Clostridium_citroniae	-26.25	3.54	1.3E-13	2.2E-12
Acidaminococcus_unclassified	-26.09	3.54	1.8E-13	2.6E-12
Clostridium_symbiosum	-26.04	3.54	2.0E-13	2.6E-12
Mitsuokella_unclassified	-25.00	3.54	1.7E-12	2.1E-11
Bacteroides_clarus	-24.80	3.54	2.6E-12	2.8E-11
Desulfovibrio_desulfuricans	-23.56	3.54	3.0E-11	3.0E-10
Olsenella_unclassified	-22.09	3.54	4.6E-10	4.2E-09
Clostridium_hathewayi	-19.35	2.28	2.0E-17	6.1E-16
Anaerotruncus_colihominis	30.00	3.52	1.7E-17	6.1E-16
Clostridium_nexile	30.00	3.52	1.7E-17	6.1E-16
AS T3 & AS T0				
Alistipes_unclassified	-32.11	7.53	2.0E-05	2.2E-04
Lachnospiraceae_bacterium_2_1_58FAA	-31.16	6.22	5.5E-07	1.3E-05
Odoribacter_splanchnicus	-30.14	7.53	6.3E-05	6.2E-04
Ruminococcus_bromii	-29.80	4.83	6.7E-10	2.7E-08
Paraprevotella_clara	-29.66	6.60	6.9E-06	8.2E-05
Paraprevotella_unclassified	-29.64	6.06	1.0E-06	1.6E-05
Alistipes_indistinctus	-29.57	4.93	1.9E-09	5.8E-08
Bacteroides_cellulosilyticus	-29.55	6.05	1.1E-06	1.6E-05
Flavonifractor_plautii	-29.28	5.91	7.2E-07	1.4E-05
Bilophila_wadsworthia	-29.02	4.00	3.8E-13	2.3E-11
Odoribacter_unclassified	-28.95	7.53	1.2E-04	1.0E-03
Clostridium_citroniae	-28.63	7.53	1.4E-04	1.1E-03
Eubacterium_ventriosum	-28.22	3.43	2.0E-16	2.3E-14
Eggerthella_lenta	-28.15	5.91	1.9E-06	2.5E-05
Pseudoflavonifractor_capillosus	-27.94	7.53	2.1E-04	1.4E-03
Anaerotruncus_colihominis	-27.79	7.53	2.2E-04	1.4E-03
Burkholderiales_bacterium_1_1_47	-27.74	7.51	2.2E-04	1.4E-03
Bifidobacterium_animalis	-26.37	7.24	2.7E-04	1.6E-03
Bacteroides_stercoris	-22.08	5.99	2.3E-04	1.4E-03
Ruminococcaceae_bacterium_D16	-21.69	7.53	4.0E-03	2.3E-02
Lactococcus_lactis	29.03	7.39	8.5E-05	7.8E-04
AS T2 & AS T0				
Odoribacter_unclassified	-29.85	6.12	1.1E-06	6.4E-05
Bifidobacterium_animalis	-26.37	5.88	7.3E-06	2.9E-04
Ruminococcaceae_bacterium_D16	-23.17	6.12	1.5E-04	3.7E-03
Eubacterium_biforme	24.87	5.75	1.5E-05	4.5E-04
Coprobacter_fastidiosus	31.45	6.12	2.8E-07	3.3E-05
AS T3 & AS T2				
Bacteroides_cellulosilyticus	-32.28	6.05	9.7E-08	2.3E-06
Alistipes_unclassified	-32.23	7.53	1.9E-05	1.9E-04
Odoribacter_splanchnicus	-31.61	7.53	2.7E-05	2.5E-04
Paraprevotella_unclassified	-30.96	6.06	3.3E-07	6.5E-06
Paraprevotella_clara	-30.82	6.60	3.0E-06	3.5E-05
Alistipes_indistinctus	-30.38	4.93	7.0E-10	2.8E-08
Clostridium_citroniae	-30.28	7.53	5.8E-05	4.9E-04
Eggerthella_lenta	-29.89	5.91	4.2E-07	7.0E-06
Lachnospiraceae_bacterium_2_1_58FAA	-29.83	6.22	1.6E-06	2.2E-05
Flavonifractor_plautii	-29.77	5.91	4.7E-07	7.0E-06

Bilophila_wadsworthia	**-28.48**	4.00	1.0E-12	6.1E-11
Coprobacter_fastidiosus	**-28.24**	7.53	1.8E-04	1.3E-03
Pseudoflavonifractor_capillosus	**-27.82**	7.53	2.2E-04	1.5E-03
Anaerotruncus_colihominis	**-27.77**	7.53	2.3E-04	1.5E-03
Ruminococcus_bromii	**-27.63**	4.83	1.0E-08	3.1E-07
Bacteroides_stercoris	**-27.54**	5.99	4.3E-06	4.6E-05
Burkholderiales_bacterium_1_1_47	**-27.42**	7.51	2.6E-04	1.6E-03
Eubacterium_ventriosum	**-27.21**	3.43	2.2E-15	2.6E-13
Eubacterium_biforme	**-24.87**	7.07	4.4E-04	2.6E-03
Oscillibacter_unclassified	**-6.50**	2.29	4.5E-03	2.5E-02
Lactococcus_lactis	**29.17**	7.39	7.9E-05	6.3E-04
AS T0 & AS				
Bacteroides_coprocola	**-30.00**	5.15	5.8E-09	8.4E-08
Bacteroides_massiliensis	**-30.00**	4.23	1.2E-12	7.4E-11
Bifidobacterium_bifidum	**-30.00**	5.12	4.6E-09	8.4E-08
Eubacterium_biforme	**-30.00**	4.84	5.7E-10	1.7E-08
Megasphaera_unclassified	**-30.00**	5.15	5.8E-09	8.4E-08
Bacteroides_faecis	**-30.00**	5.15	5.8E-09	8.4E-08
Coprobacter_fastidiosus	**-29.91**	5.15	6.4E-09	8.4E-08
Bacteroides_finegoldii	**-29.77**	5.15	7.5E-09	9.0E-08
Lactococcus_lactis	**-28.44**	5.10	2.5E-08	2.7E-07
Bacteroides_salyersiae	**-22.85**	5.15	9.2E-06	7.3E-05
Odoribacter_unclassified	**26.09**	5.09	3.0E-07	2.5E-06
Clostridium_citroniae	**27.08**	5.09	1.0E-07	9.5E-07
Clostridium_hathewayi	**27.69**	3.23	9.9E-18	1.2E-15
Clostridium_symbiosum	**28.15**	5.09	3.2E-08	3.2E-07
Collinsella_intestinalis	**30.00**	4.53	3.6E-11	1.4E-09
AS T2 & AS				
Bacteroides_coprocola	**-30.00**	5.15	5.8E-09	1.0E-07
Bacteroides_massiliensis	**-30.00**	4.23	1.2E-12	7.4E-11
Bifidobacterium_animalis	**-30.00**	4.95	1.4E-09	5.5E-08
Bifidobacterium_bifidum	**-30.00**	5.12	4.6E-09	1.0E-07
Megasphaera_unclassified	**-30.00**	5.15	5.8E-09	1.0E-07
Bacteroides_faecis	**-29.95**	5.15	6.1E-09	1.0E-07
Bacteroides_finegoldii	**-29.83**	5.15	7.0E-09	1.0E-07
Ruminococcaceae_bacterium_D16	**-28.96**	5.15	1.9E-08	2.2E-07
Lactococcus_lactis	**-28.58**	5.10	2.1E-08	2.3E-07
Bacteroides_salyersiae	**-23.05**	5.15	7.7E-06	6.5E-05
Collinsella_intestinalis	**24.85**	4.53	4.2E-08	4.2E-07
Clostridium_symbiosum	**27.49**	5.09	6.6E-08	6.1E-07
Clostridium_citroniae	**28.73**	5.09	1.7E-08	2.2E-07
Clostridium_hathewayi	**29.71**	3.23	3.6E-20	4.3E-18
AS T3 & AS				
Alistipes_unclassified	**-30.00**	6.77	9.3E-06	5.3E-05
Bacteroides_cellulosilyticus	**-30.00**	5.45	3.8E-08	5.2E-07
Bacteroides_coprocola	**-30.00**	6.77	9.3E-06	5.3E-05
Bacteroides_massiliensis	**-30.00**	5.56	6.8E-08	8.1E-07
Bacteroides_stercoris	**-30.00**	5.39	2.7E-08	5.0E-07
Bifidobacterium_animalis	**-30.00**	6.51	4.0E-06	3.0E-05
Bifidobacterium_bifidum	**-30.00**	6.72	8.2E-06	5.3E-05
Eubacterium_biforme	**-30.00**	6.36	2.4E-06	2.0E-05
Eubacterium_ventriosum	**-30.00**	3.14	1.1E-21	1.3E-19
Lachnospiraceae_bacterium_2_1_58FAA	**-30.00**	5.60	8.6E-08	9.3E-07
Megasphaera_unclassified	**-30.00**	6.77	9.3E-06	5.3E-05
Odoribacter_splanchnicus	**-30.00**	6.77	9.3E-06	5.3E-05
Paraprevotella_clara	**-30.00**	5.93	4.3E-07	4.3E-06
Paraprevotella_unclassified	**-30.00**	5.46	3.9E-08	5.2E-07

Ruminococcus_bromii	-30.00	4.37	6.4E-12	2.5E-10
Anaerotruncus_colihominis	-29.80	6.77	1.1E-05	5.8E-05
Eggerthella_lenta	-29.53	5.33	2.9E-08	5.0E-07
Bacteroides_faecis	-29.45	6.77	1.4E-05	7.0E-05
Alistipes_indistinctus	-28.91	4.45	8.5E-11	2.5E-09
Bilophila_wadsworthia	-27.55	3.63	3.3E-14	2.0E-12
Ruminococcaceae_bacterium_D16	-27.48	6.77	4.9E-05	2.3E-04
Coprobacter_fastidiosus	-26.70	6.77	8.0E-05	3.6E-04
Bacteroides_finegoldii	-26.44	6.77	9.3E-05	4.1E-04
Burkholderiales_bacterium_1_1_47	-26.09	6.75	1.1E-04	4.7E-04
Flavonifractor_plautii	-25.05	5.32	2.5E-06	2.0E-05
Bacteroides_salyersiae	-24.92	6.77	2.3E-04	9.5E-04
Pseudoflavonifractor_capillosus	-24.01	6.77	3.9E-04	1.5E-03
Parabacteroides_distasonis	-7.37	2.70	6.3E-03	2.4E-02
Clostridium_hathewayi	26.70	4.19	1.9E-10	4.5E-09
Clostridium_symbiosum	27.05	6.64	4.6E-05	2.3E-04
Collinsella_intestinalis	28.63	5.91	1.3E-06	1.2E-05

Significance is defined by P-adjusted ≤ 0.05. Red indicates species with reduced abundance while green represents species with increased abundance in that pairwise comparison.

C) Significant pairwise comparisons of psoriatic arthritis (PSA) time-point, established samples and controls T0

PSA T0 & Control T0	**log2FoldChange**	**lfcSE**	**pvalue**	**padj**
Bacteroides_finegoldii	-30.00	4.59	6.6E-11	2.6E-09
Bacteroides_salyersiae	-30.00	4.59	6.6E-11	2.6E-09
Lactococcus_lactis	-29.84	4.04	1.5E-13	1.8E-11
Acidaminococcus_unclassified	-29.74	4.59	9.7E-11	2.6E-09
Mitsuokella_unclassified	-29.65	4.59	1.1E-10	2.6E-09
Mitsuokella_multacida	-27.40	4.59	2.5E-09	4.9E-08
Turicibacter_unclassified	-26.34	4.59	9.9E-09	1.7E-07
Oxalobacter_formigenes	-24.09	4.59	1.6E-07	2.0E-06
Bacteroides_clarus	-24.04	4.59	1.7E-07	2.0E-06
Ruminococcus_torques	2.44	0.62	8.2E-05	8.8E-04
Clostridium_asparagiforme	13.32	4.22	1.6E-03	1.6E-02
Erysipelotrichaceae_bacterium_21_3	24.77	4.56	5.5E-08	8.1E-07
PSA T2 & Control T0				
Bacteroides_salyersiae	-30.00	5.68	1.3E-07	2.3E-06
Desulfovibrio_piger	-30.00	5.02	2.2E-09	2.6E-07
Megamonas_unclassified	-30.00	5.68	1.3E-07	2.3E-06
Phascolarctobacterium_succinatutens	-30.00	5.68	1.3E-07	2.3E-06
Megamonas_hypermegale	-30.00	5.68	1.3E-07	2.3E-06
Alistipes_unclassified	-29.93	5.68	1.4E-07	2.3E-06
Bacteroides_finegoldii	-29.80	5.68	1.5E-07	2.3E-06
Mitsuokella_unclassified	-28.84	5.68	3.8E-07	4.5E-06
Mitsuokella_multacida	-26.97	5.68	2.1E-06	1.9E-05
Bacteroides_faecis	-26.92	5.64	1.8E-06	1.8E-05
Lactococcus_lactis	-26.29	5.00	1.4E-07	2.3E-06
Streptococcus_vestibularis	-26.17	5.68	4.1E-06	3.0E-05
Acidaminococcus_unclassified	-24.04	5.68	2.3E-05	1.6E-04
Lachnospiraceae_bacterium_1_4_56FAA	-23.87	5.13	3.3E-06	2.6E-05
Bacteroides_clarus	-23.20	5.68	4.4E-05	2.9E-04
Oxalobacter_formigenes	-22.08	5.68	1.0E-04	6.3E-04
Clostridium_asparagiforme	14.76	5.19	4.4E-03	2.6E-02
Erysipelotrichaceae_bacterium_21_3	26.40	5.61	2.5E-06	2.1E-05
Anaerotruncus_colihominis	27.75	5.61	7.6E-07	8.1E-06
Turicibacter_sanguinis	28.74	5.61	3.0E-07	4.0E-06

PSA T3 & Control T0				
Bacteroides_stercoris	**-30.00**	4.54	3.8E-11	2.2E-09
Phascolarctobacterium_succinatutens	**-30.00**	6.40	2.7E-06	4.0E-05
Roseburia_intestinalis	**-30.00**	3.10	3.5E-22	4.1E-20
Megamonas_unclassified	**-29.72**	6.40	3.4E-06	4.5E-05
Bacteroides_salyersiae	**-29.20**	6.40	5.0E-06	5.9E-05
Desulfovibrio_piger	**-29.04**	5.65	2.8E-07	8.2E-06
Megamonas_hypermegale	**-28.46**	6.40	8.6E-06	9.3E-05
Bacteroides_finegoldii	**-28.16**	6.40	1.1E-05	1.1E-04
Alistipes_unclassified	**-27.96**	6.40	1.2E-05	1.1E-04
Paraprevotella_xylaniphila	**-27.84**	5.30	1.5E-07	6.0E-06
Mitsuokella_unclassified	**-27.22**	6.40	2.1E-05	1.8E-04
Lactococcus_lactis	**-27.01**	5.63	1.6E-06	3.2E-05
Bacteroides_faecis	**-25.90**	6.35	4.6E-05	3.6E-04
Mitsuokella_multacida	**-25.66**	6.40	6.0E-05	4.5E-04
Bacteroides_clarus	**-24.76**	6.40	1.1E-04	6.9E-04
Desulfovibrio_desulfuricans	**-24.73**	6.40	1.1E-04	6.9E-04
Acidaminococcus_unclassified	**-24.46**	6.40	1.3E-04	7.8E-04
Bacteroides_fragilis	**-23.59**	6.09	1.1E-04	6.9E-04
Flavonifractor_plautii	**-22.57**	4.46	4.1E-07	9.6E-06
Turicibacter_unclassified	**-22.49**	6.40	4.4E-04	2.4E-03
Streptococcus_vestibularis	**-21.43**	6.40	8.1E-04	4.2E-03
Lachnospiraceae_bacterium_1_4_56FAA	**-20.87**	5.78	3.0E-04	1.7E-03
Oxalobacter_formigenes	**-20.28**	6.40	1.5E-03	7.5E-03
Erysipelotrichaceae_bacterium_21_3	**30.00**	6.31	2.0E-06	3.3E-05
PSA & Control T0				
Mitsuokella_unclassified	**-30.00**	3.47	5.1E-18	2.4E-16
Mitsuokella_multacida	**-29.95**	3.47	5.8E-18	2.4E-16
Bacteroides_clarus	**-29.93**	3.47	6.0E-18	2.4E-16
Veillonella_dispar	**-26.44**	3.47	2.4E-14	4.8E-13
Desulfovibrio_piger	**-14.10**	3.04	3.4E-06	4.0E-05
Clostridium_asparagiforme	**13.60**	3.21	2.3E-05	2.4E-04
Erysipelotrichaceae_bacterium_21_3	**22.65**	3.47	6.8E-11	8.9E-10
Actinomyces_graevenitzii	**25.33**	3.47	3.0E-13	4.4E-12
Eubacterium_cylindroides	**25.65**	3.47	1.5E-13	2.5E-12
Turicibacter_sanguinis	**27.81**	3.47	1.1E-15	2.7E-14
Anaerotruncus_colihominis	**29.33**	3.47	3.0E-17	8.8E-16
PSA T3 & PSA T0				
Roseburia_intestinalis	**-31.41**	3.36	7.9E-21	9.3E-19
Megamonas_unclassified	**-31.28**	7.00	7.8E-06	1.8E-04
Alistipes_unclassified	**-30.26**	7.00	1.5E-05	2.6E-04
Phascolarctobacterium_succinatutens	**-29.79**	7.00	2.1E-05	2.7E-04
Megamonas_hypermegale	**-29.65**	7.00	2.3E-05	2.7E-04
Bacteroides_fragilis	**-28.65**	6.66	1.7E-05	2.6E-04
Bacteroides_stercoris	**-28.57**	4.95	7.7E-09	4.6E-07
Bacteroides_faecis	**-27.88**	6.95	6.0E-05	6.4E-04
Clostridium_asparagiforme	**-27.85**	6.49	1.8E-05	2.6E-04
Desulfovibrio_piger	**-27.82**	6.18	6.7E-06	1.8E-04
Streptococcus_vestibularis	**-26.66**	7.00	1.4E-04	1.4E-03
Flavonifractor_plautii	**-25.08**	4.86	2.4E-07	9.6E-06
Desulfovibrio_desulfuricans	**-24.60**	7.00	4.4E-04	3.7E-03
Lachnospiraceae_bacterium_1_4_56FAA	**-23.28**	6.32	2.3E-04	2.1E-03
Paraprevotella_xylaniphila	**-18.53**	5.80	1.4E-03	1.1E-02
PSA T2 & PSA T0				
Alistipes_unclassified	**-32.23**	6.35	3.8E-07	1.1E-05
Megamonas_unclassified	**-31.57**	6.35	6.6E-07	1.5E-05
Streptococcus_vestibularis	**-31.40**	6.35	7.5E-07	1.5E-05

Megamonas_hypermegale	**-31.19**	6.35	8.9E-07	1.5E-05
Phascolarctobacterium_succinatutens	**-29.79**	6.35	2.7E-06	4.0E-05
Bacteroides_faecis	**-28.90**	6.30	4.5E-06	5.4E-05
Desulfovibrio_piger	**-28.78**	5.60	2.8E-07	1.1E-05
Lachnospiraceae_bacterium_1_4_56FAA	**-26.27**	5.73	4.5E-06	5.4E-05
Turicibacter_unclassified	**28.77**	6.31	5.2E-06	5.5E-05
Anaerotruncus_colihominis	**34.25**	6.31	5.8E-08	3.4E-06
Turicibacter_sanguinis	**38.30**	6.31	1.3E-09	1.5E-07
PSA T3 & PSA T2				
Turicibacter_sanguinis	**-31.84**	7.72	3.8E-05	9.1E-04
Bacteroides_fragilis	**-31.59**	7.35	1.7E-05	6.8E-04
Anaerotruncus_colihominis	**-31.47**	7.72	4.6E-05	9.1E-04
Roseburia_intestinalis	**-30.80**	3.67	5.0E-17	5.9E-15
Clostridium_asparagiforme	**-29.29**	7.16	4.3E-05	9.1E-04
Flavonifractor_plautii	**-26.99**	5.35	4.5E-07	2.7E-05
Desulfovibrio_desulfuricans	**-26.75**	7.72	5.3E-04	7.9E-03
Turicibacter_unclassified	**-24.92**	7.72	1.3E-03	1.6E-02
Bacteroides_stercoris	**-21.56**	5.45	7.6E-05	1.3E-03
Paraprevotella_xylaniphila	**-20.47**	6.39	1.4E-03	1.6E-02
PSA T0 & PSA				
Turicibacter_sanguinis	**-37.37**	4.52	1.3E-16	1.6E-14
Anaerotruncus_colihominis	**-35.82**	4.52	2.2E-15	1.3E-13
Acidaminococcus_unclassified	**-33.82**	4.52	7.2E-14	2.1E-12
Eubacterium_cylindroides	**-32.31**	4.52	8.6E-13	2.0E-11
Actinomyces_graevenitzii	**-30.54**	4.52	1.4E-11	2.7E-10
Turicibacter_unclassified	**-30.00**	4.52	3.1E-11	5.3E-10
Lactococcus_lactis	**-29.90**	3.97	5.3E-14	2.1E-12
Bacteroides_finegoldii	**-29.06**	4.52	1.3E-10	1.9E-09
Oxalobacter_formigenes	**-27.95**	4.52	6.2E-10	8.1E-09
Bacteroides_salyersiae	**-25.01**	4.52	3.1E-08	3.3E-07
Desulfovibrio_piger	**12.87**	3.93	1.0E-03	1.0E-02
Veillonella_dispar	**26.20**	4.48	4.9E-09	5.8E-08
PSA T2 & PSA				
Megamonas_unclassified	**-32.57**	5.62	6.7E-09	4.0E-07
Megamonas_hypermegale	**-32.09**	5.62	1.1E-08	4.4E-07
Lachnospiraceae_bacterium_1_4_56FAA	**-31.16**	5.08	8.3E-10	9.8E-08
Bacteroides_faecis	**-31.08**	5.58	2.5E-08	7.4E-07
Alistipes_unclassified	**-30.14**	5.62	8.1E-08	1.7E-06
Phascolarctobacterium_succinatutens	**-29.22**	5.62	2.0E-07	2.9E-06
Bacteroides_finegoldii	**-28.86**	5.62	2.8E-07	3.7E-06
Streptococcus_vestibularis	**-28.62**	5.62	3.5E-07	4.1E-06
Acidaminococcus_unclassified	**-28.12**	5.62	5.6E-07	6.0E-06
Eubacterium_cylindroides	**-27.55**	5.62	9.4E-07	9.3E-06
Lactococcus_lactis	**-26.34**	4.94	9.9E-08	1.7E-06
Actinomyces_graevenitzii	**-26.26**	5.62	3.0E-06	2.7E-05
Oxalobacter_formigenes	**-25.95**	5.62	3.9E-06	3.3E-05
Bacteroides_salyersiae	**-25.01**	5.62	8.5E-06	6.7E-05
Desulfovibrio_piger	**-15.90**	4.97	1.4E-03	1.0E-02
Veillonella_dispar	**29.60**	5.55	9.4E-08	1.7E-06
PSA T3 & PSA				
Anaerotruncus_colihominis	**-33.04**	6.34	1.9E-07	5.6E-06
Megamonas_unclassified	**-32.29**	6.34	3.6E-07	8.5E-06
Turicibacter_sanguinis	**-30.91**	6.34	1.1E-06	1.6E-05
Megamonas_hypermegale	**-30.55**	6.34	1.5E-06	1.7E-05
Bacteroides_faecis	**-30.06**	6.30	1.8E-06	1.8E-05
Roseburia_intestinalis	**-29.71**	3.07	4.2E-22	5.0E-20

Bacteroides_fragilis	**-29.46**	6.04	1.1E-06	1.6E-05
Phascolarctobacterium_succinatutens	**-29.22**	6.34	4.1E-06	3.5E-05
Bacteroides_stercoris	**-29.07**	4.50	1.0E-10	6.1E-09
Acidaminococcus_unclassified	**-28.54**	6.34	6.8E-06	5.4E-05
Alistipes_unclassified	**-28.18**	6.34	8.9E-06	6.2E-05
Lachnospiraceae_bacterium_1_4_56FAA	**-28.16**	5.73	8.9E-07	1.6E-05
Clostridium_asparagiforme	**-28.13**	5.88	1.7E-06	1.8E-05
Flavonifractor_plautii	**-27.62**	4.42	4.1E-10	1.6E-08
Eubacterium_cylindroides	**-27.41**	6.34	1.5E-05	1.0E-04
Bacteroides_finegoldii	**-27.22**	6.34	1.8E-05	1.1E-04
Lactococcus_lactis	**-27.06**	5.58	1.3E-06	1.6E-05
Actinomyces_graevenitzii	**-26.25**	6.34	3.5E-05	2.1E-04
Turicibacter_unclassified	**-26.15**	6.34	3.7E-05	2.1E-04
Desulfovibrio_desulfuricans	**-26.10**	6.34	3.9E-05	2.1E-04
Bacteroides_salyersiae	**-24.21**	6.34	1.4E-04	7.0E-04
Oxalobacter_formigenes	**-24.14**	6.34	1.4E-04	7.0E-04
Streptococcus_vestibularis	**-23.88**	6.34	1.7E-04	7.9E-04
Paraprevotella_xylaniphila	**-23.47**	5.26	8.1E-06	6.0E-05
Desulfovibrio_piger	**-14.94**	5.61	7.8E-03	3.5E-02
Veillonella_dispar	**29.57**	6.25	2.2E-06	2.0E-05

Significance is defined by P-adjusted ≤ 0.05. Red colour indicates species with reduced abundance while green represents species with increased abundance in that pairwise comparison.

D) Significant pairwise comparisons of Inflammatory bowel diseases (IBD) time-point, established samples and controls T0

IBD T0 & Control T0	log2FoldChange	lfcSE	pvalue	padj
Adlercreutzia_equolifaciens	**-30.00**	2.73	4.4E-28	2.5E-26
Akkermansia_muciniphila	**-30.00**	3.22	1.2E-20	2.0E-19
Alistipes_finegoldii	**-30.00**	3.12	7.9E-22	1.5E-20
Alistipes_onderdonkii	**-30.00**	3.49	8.0E-18	7.0E-17
Alistipes_senegalensis	**-30.00**	4.01	6.9E-14	4.4E-13
Bacteroidales_bacterium_ph8	**-30.00**	3.07	1.5E-22	3.5E-21
Bacteroides_caccae	**-30.00**	3.53	1.8E-17	1.5E-16
Bacteroides_cellulosilyticus	**-30.00**	4.26	2.0E-12	5.5E-12
Bacteroides_coprocola	**-30.00**	4.26	2.0E-12	5.5E-12
Bacteroides_dorei	**-30.00**	3.46	4.6E-18	4.4E-17
Bacteroides_finegoldii	**-30.00**	4.26	2.0E-12	5.5E-12
Bacteroides_massiliensis	**-30.00**	3.74	1.0E-15	6.7E-15
Bacteroides_plebeius	**-30.00**	4.26	2.0E-12	5.5E-12
Bacteroides_salyersiae	**-30.00**	4.26	2.0E-12	5.5E-12
Bifidobacterium_catenulatum	**-30.00**	4.26	2.0E-12	5.5E-12
Clostridium_leptum	**-30.00**	3.28	5.2E-20	5.9E-19
Coprococcus_sp_ART55_1	**-30.00**	4.26	2.0E-12	5.5E-12
Desulfovibrio_piger	**-30.00**	4.26	2.0E-12	5.5E-12
Eubacterium_biforme	**-30.00**	4.26	2.0E-12	5.5E-12
Eubacterium_siraeum	**-30.00**	3.02	3.1E-23	8.9E-22
Odoribacter_splanchnicus	**-30.00**	4.26	2.0E-12	5.5E-12
Parabacteroides_unclassified	**-30.00**	4.26	2.0E-12	5.5E-12
Paraprevotella_clara	**-30.00**	4.26	2.0E-12	5.5E-12
Paraprevotella_unclassified	**-30.00**	4.12	3.2E-13	1.7E-12
Prevotella_copri	**-30.00**	4.26	2.0E-12	5.5E-12
Roseburia_inulinivorans	**-30.00**	2.60	1.1E-30	1.2E-28
Roseburia_unclassified	**-30.00**	4.26	2.0E-12	5.5E-12
Ruminococcus_bromii	**-30.00**	3.27	4.4E-20	5.9E-19
Ruminococcus_lactaris	**-30.00**	4.04	1.1E-13	6.6E-13
Coprobacter_fastidiosus	**-29.95**	4.26	2.2E-12	5.9E-12
Bilophila_wadsworthia	**-29.80**	2.95	4.5E-24	1.7E-22
Mitsuokella_unclassified	**-29.36**	4.26	5.8E-12	1.5E-11
Acidaminococcus_unclassified	**-28.96**	4.26	1.1E-11	2.8E-11
Odoribacter_unclassified	**-28.77**	4.26	1.5E-11	3.8E-11
Ruminococcus_callidus	**-28.75**	4.26	1.6E-11	3.9E-11
Mitsuokella_multacida	**-27.64**	4.26	9.0E-11	2.2E-10
Bacteroides_eggerthii	**-27.35**	4.26	1.4E-10	3.4E-10
Bacteroides_clarus	**-25.83**	4.26	1.4E-09	3.1E-09
Paraprevotella_xylaniphila	**-25.72**	4.26	1.6E-09	3.6E-09
Megamonas_unclassified	**-24.91**	4.26	5.1E-09	1.1E-08
Desulfovibrio_desulfuricans	**-24.71**	4.26	6.8E-09	1.5E-08
Bacteroides_faecis	**-23.68**	4.26	2.8E-08	5.8E-08
Lachnospiraceae_bacterium_3_1_57FAA_CT1	**-22.98**	4.26	7.0E-08	1.4E-07
Pseudoflavonifractor_capillosus	**-22.62**	4.26	1.1E-07	2.3E-07
Olsenella_unclassified	**-21.93**	4.27	2.7E-07	5.4E-07
Roseburia_hominis	**-18.16**	1.98	4.7E-20	5.9E-19
Alistipes_indistinctus	**-15.13**	2.62	7.9E-09	1.7E-08
Barnesiella_intestinihominis	**-10.82**	2.81	1.2E-04	2.3E-04
Bilophila_unclassified	**-8.08**	2.26	3.5E-04	6.7E-04
Bacteroides_thetaiotaomicron	**-7.63**	3.14	1.5E-02	2.6E-02
Eubacterium_rectale	**-5.54**	2.42	2.2E-02	3.8E-02
Alistipes_shahii	**-5.45**	2.40	2.3E-02	4.0E-02
Subdoligranulum_unclassified	**-5.38**	1.69	1.5E-03	2.7E-03
Ruminococcus_gnavus	**8.55**	2.52	6.8E-04	1.2E-03
Veillonella_parvula	**8.63**	3.88	2.6E-02	4.3E-02
Enterococcus_faecalis	**9.33**	4.14	2.4E-02	4.1E-02
Veillonella_unclassified	**10.00**	2.98	7.8E-04	1.4E-03

Bacteroides_fragilis	**14.11**	4.14	6.5E-04	1.2E-03
Lachnospiraceae_bacterium_4_1_37FAA	**25.48**	4.17	1.0E-09	2.4E-09
Clostridium_nexile	**29.94**	3.45	4.0E-18	4.2E-17
Anaerostipes_unclassified	**30.00**	4.17	6.5E-13	2.8E-12
Blautia_hansenii	**30.00**	4.17	6.5E-13	2.8E-12
Blautia_producta	**30.00**	4.06	1.5E-13	8.8E-13
Clostridiales_bacterium_1_7_47FAA	**30.00**	4.17	6.5E-13	2.8E-12
Dorea_unclassified	**30.00**	4.17	6.5E-13	2.8E-12
Lachnospiraceae_bacterium_6_1_63FAA	**30.00**	3.58	5.5E-17	4.2E-16
Lachnospiraceae_bacterium_9_1_43BFAA	**30.00**	3.63	1.3E-16	9.2E-16
Morganella_morganii	**30.00**	4.17	6.5E-13	2.8E-12
Streptococcus_gordonii	**30.00**	4.17	6.5E-13	2.8E-12
IBD T2 & Control T0				
Akkermansia_muciniphila	**-30.00**	3.22	1.2E-20	1.8E-19
Alistipes_finegoldii	**-30.00**	3.12	7.9E-22	1.3E-20
Alistipes_onderdonkii	**-30.00**	3.49	8.0E-18	1.0E-16
Alistipes_senegalensis	**-30.00**	4.01	6.9E-14	5.6E-13
Alistipes_shahii	**-30.00**	2.61	1.4E-30	5.5E-29
Bacteroidales_bacterium_ph8	**-30.00**	3.07	1.5E-22	2.9E-21
Bacteroides_caccae	**-30.00**	3.53	1.8E-17	2.0E-16
Bacteroides_cellulosilyticus	**-30.00**	4.26	2.0E-12	6.4E-12
Bacteroides_coprocola	**-30.00**	4.26	2.0E-12	6.4E-12
Bacteroides_faecis	**-30.00**	4.26	2.0E-12	6.4E-12
Bacteroides_finegoldii	**-30.00**	4.26	2.0E-12	6.4E-12
Bacteroides_massiliensis	**-30.00**	3.74	1.0E-15	9.6E-15
Bacteroides_plebeius	**-30.00**	4.26	2.0E-12	6.4E-12
Bacteroides_salyersiae	**-30.00**	4.26	2.0E-12	6.4E-12
Bifidobacterium_catenulatum	**-30.00**	4.26	2.0E-12	6.4E-12
Coprobacter_fastidiosus	**-30.00**	4.26	2.0E-12	6.4E-12
Coprococcus_sp_ART55_1	**-30.00**	4.26	2.0E-12	6.4E-12
Desulfovibrio_piger	**-30.00**	4.26	2.0E-12	6.4E-12
Eubacterium_biforme	**-30.00**	4.26	2.0E-12	6.4E-12
Eubacterium_eligens	**-30.00**	2.56	1.2E-31	1.4E-29
Eubacterium_siraeum	**-30.00**	3.02	3.1E-23	7.6E-22
Odoribacter_splanchnicus	**-30.00**	4.26	2.0E-12	6.4E-12
Odoribacter_unclassified	**-30.00**	4.26	2.0E-12	6.4E-12
Parabacteroides_unclassified	**-30.00**	4.26	2.0E-12	6.4E-12
Paraprevotella_clara	**-30.00**	4.26	2.0E-12	6.4E-12
Paraprevotella_unclassified	**-30.00**	4.12	3.2E-13	2.3E-12
Prevotella_copri	**-30.00**	4.26	2.0E-12	6.4E-12
Roseburia_inulinivorans	**-30.00**	2.60	1.1E-30	5.5E-29
Roseburia_unclassified	**-30.00**	4.26	2.0E-12	6.4E-12
Ruminococcus_lactaris	**-30.00**	4.04	1.1E-13	8.3E-13
Mitsuokella_unclassified	**-29.51**	4.26	4.5E-12	1.4E-11
Ruminococcus_callidus	**-29.14**	4.26	8.2E-12	2.4E-11
Mitsuokella_multacida	**-28.04**	4.26	4.8E-11	1.3E-10
Bacteroides_eggerthii	**-27.65**	4.26	8.9E-11	2.3E-10
Paraprevotella_xylaniphila	**-26.81**	4.26	3.2E-10	7.8E-10
Bacteroides_clarus	**-26.40**	4.26	6.0E-10	1.4E-09
Desulfovibrio_desulfuricans	**-25.72**	4.26	1.6E-09	3.7E-09
Megamonas_unclassified	**-25.22**	4.26	3.3E-09	7.4E-09
Olsenella_unclassified	**-24.94**	4.27	5.0E-09	1.1E-08
Pseudoflavonifractor_capillosus	**-23.43**	4.26	3.9E-08	8.2E-08
Lachnospiraceae_bacterium_3_1_57FAA_CT1	**-23.27**	4.26	4.8E-08	1.0E-07
Roseburia_hominis	**-20.25**	2.04	3.3E-23	7.6E-22
Alistipes_indistinctus	**-17.29**	2.67	8.8E-11	2.3E-10
Alistipes_putredinis	**-7.54**	3.22	1.9E-02	3.8E-02
Ruminococcus_gnavus	**8.45**	2.52	7.8E-04	1.6E-03
Bacteroides_fragilis	**11.50**	4.14	5.5E-03	1.1E-02

Lachnospiraceae_bacterium_4_1_37FAA	**23.52**	4.17	1.7E-08	3.7E-08
Lachnospiraceae_bacterium_6_1_63FAA	**23.71**	3.58	3.6E-11	1.0E-10
Blautia_producta	**25.31**	4.06	4.7E-10	1.1E-09
Clostridiales_bacterium_1_7_47FAA	**26.92**	4.17	1.1E-10	2.7E-10
Blautia_hansenii	**27.14**	4.17	7.9E-11	2.1E-10
Clostridium_nexile	**27.93**	3.45	5.8E-16	6.0E-15
Anaerostipes_unclassified	**28.32**	4.17	1.1E-11	3.3E-11
Streptococcus_gordonii	**28.81**	4.17	5.0E-12	1.5E-11
Morganella_morganii	**28.82**	4.17	5.0E-12	1.5E-11
Lachnospiraceae_bacterium_9_1_43BFAA	**28.91**	3.63	1.6E-15	1.4E-14
Dorea_unclassified	**29.38**	4.17	1.9E-12	6.4E-12
Lachnospiraceae_bacterium_5_1_57FAA	**30.00**	4.17	6.5E-13	4.4E-12
IBD T3 & Control T0				
Akkermansia_muciniphila	**-30.00**	3.63	1.4E-16	4.0E-15
Bacteroidales_bacterium_ph8	**-30.00**	3.46	4.5E-18	1.8E-16
Bacteroides_caccae	**-30.00**	3.97	4.3E-14	7.0E-13
Bacteroides_cellulosilyticus	**-30.00**	4.80	4.1E-10	1.9E-09
Bacteroides_coprocola	**-30.00**	4.80	4.1E-10	1.9E-09
Bacteroides_finegoldii	**-30.00**	4.80	4.1E-10	1.9E-09
Bacteroides_massiliensis	**-30.00**	4.21	1.0E-12	1.3E-11
Bacteroides_plebeius	**-30.00**	4.80	4.1E-10	1.9E-09
Bacteroides_salyersiae	**-30.00**	4.80	4.1E-10	1.9E-09
Clostridium_leptum	**-30.00**	3.69	4.4E-16	9.9E-15
Coprococcus_sp_ART55_1	**-30.00**	4.80	4.1E-10	1.9E-09
Desulfovibrio_piger	**-30.00**	4.80	4.1E-10	1.9E-09
Eubacterium_biforme	**-30.00**	4.80	4.1E-10	1.9E-09
Eubacterium_siraeum	**-30.00**	3.41	1.3E-18	1.5E-16
Parabacteroides_unclassified	**-30.00**	4.80	4.1E-10	1.9E-09
Paraprevotella_clara	**-30.00**	4.80	4.1E-10	1.9E-09
Paraprevotella_unclassified	**-30.00**	4.63	9.6E-11	1.0E-09
Prevotella_copri	**-30.00**	4.80	4.1E-10	1.9E-09
Roseburia_unclassified	**-30.00**	4.80	4.1E-10	1.9E-09
Ruminococcus_lactaris	**-30.00**	4.55	4.2E-11	4.7E-10
Ruminococcus_callidus	**-28.77**	4.80	2.0E-09	8.6E-09
Alistipes_senegalensis	**-28.64**	4.51	2.2E-10	1.9E-09
Bifidobacterium_catenulatum	**-28.47**	4.80	3.0E-09	1.2E-08
Mitsuokella_unclassified	**-27.92**	4.80	6.0E-09	2.1E-08
Mitsuokella_multacida	**-27.74**	4.80	7.4E-09	2.6E-08
Bacteroides_eggerthii	**-27.19**	4.80	1.5E-08	4.6E-08
Paraprevotella_xylaniphila	**-25.93**	4.80	6.6E-08	2.0E-07
Bacteroides_clarus	**-25.80**	4.80	7.6E-08	2.3E-07
Coprobacter_fastidiosus	**-25.60**	4.80	9.6E-08	2.7E-07
Bacteroides_faecis	**-25.05**	4.80	1.8E-07	4.9E-07
Megamonas_unclassified	**-24.89**	4.80	2.2E-07	5.7E-07
Desulfovibrio_desulfuricans	**-23.47**	4.80	1.0E-06	2.5E-06
Odoribacter_unclassified	**-23.12**	4.80	1.4E-06	3.6E-06
Pseudoflavonifractor_capillosus	**-22.97**	4.80	1.7E-06	4.1E-06
Lachnospiraceae_bacterium_3_1_57FAA_CT1	**-22.83**	4.80	2.0E-06	4.7E-06
Roseburia_hominis	**-20.01**	2.31	4.7E-18	1.8E-16
Olsenella_unclassified	**-19.25**	4.80	6.1E-05	1.4E-04
Alistipes_indistinctus	**-17.04**	3.01	1.5E-08	4.6E-08
Alistipes_onderdonkii	**-10.72**	3.75	4.3E-03	9.3E-03
Eubacterium_rectale	**-7.44**	2.71	6.1E-03	1.3E-02
Streptococcus_thermophilus	**-7.10**	3.04	2.0E-02	4.0E-02
Subdoligranulum_unclassified	**-5.36**	1.90	4.9E-03	1.1E-02
Ruminococcus_gnavus	**9.04**	2.83	1.4E-03	3.2E-03
Bacteroides_fragilis	**11.04**	4.65	1.8E-02	3.7E-02
Enterococcus_faecalis	**13.31**	4.65	4.2E-03	9.3E-03
Lachnospiraceae_bacterium_6_1_63FAA	**20.31**	4.02	4.3E-07	1.1E-06
Blautia_producta	**24.45**	4.56	8.2E-08	2.4E-07

Clostridiales_bacterium_1_7_47FAA	**24.57**	4.68	1.5E-07	4.3E-07
Blautia_hansenii	**26.69**	4.68	1.2E-08	4.0E-08
Lachnospiraceae_bacterium_5_1_57FAA	**27.26**	4.68	5.8E-09	2.1E-08
Dorea_unclassified	**27.54**	4.68	4.0E-09	1.5E-08
Morganella_morganii	**27.71**	4.68	3.3E-09	1.3E-08
Anaerostipes_unclassified	**28.50**	4.68	1.2E-09	5.1E-09
Lachnospiraceae_bacterium_9_1_43BFAA	**29.69**	4.07	2.8E-13	4.1E-12
Clostridium_nexile	**30.00**	3.87	8.9E-15	1.7E-13
Lachnospiraceae_bacterium_4_1_37FAA	**30.00**	4.68	1.5E-10	1.4E-09
IBD T3 & IBD T0				
Streptococcus_gordonii	**-35.57**	5.74	5.8E-10	9.4E-09
Alistipes_onderdonkii	**19.28**	4.64	3.3E-05	3.4E-04
Alistipes_finegoldii	**23.42**	4.14	1.5E-08	1.9E-07
Roseburia_inulinivorans	**25.46**	3.41	8.0E-14	4.6E-12
Adlercreutzia_equolifaciens	**25.58**	3.59	9.8E-13	3.7E-11
Bacteroides_dorei	**26.81**	4.61	5.8E-09	8.3E-08
Ruminococcus_bromii	**28.61**	4.34	4.2E-11	9.6E-10
Odoribacter_splanchnicus	**29.18**	5.71	3.2E-07	3.7E-06
Bilophila_wadsworthia	**29.37**	3.89	4.1E-14	4.6E-12
Acidaminococcus_unclassified	**35.96**	5.71	3.0E-10	5.7E-09
Lachnospiraceae_bacterium_5_1_57FAA	**38.70**	5.71	1.2E-11	3.5E-10
IBD T2 & IBD T0				
Eubacterium_eligens	**-27.25**	3.13	2.9E-18	3.3E-16
Alistipes_shahii	**-24.55**	3.19	1.3E-14	3.8E-13
Bacteroides_dorei	**26.12**	4.28	1.1E-09	1.4E-08
Bilophila_wadsworthia	**26.96**	3.62	9.3E-14	2.1E-12
Clostridium_leptum	**27.02**	4.04	2.3E-11	3.3E-10
Adlercreutzia_equolifaciens	**27.68**	3.34	1.2E-16	6.9E-15
Ruminococcus_bromii	**29.37**	4.03	3.3E-13	6.3E-12
Acidaminococcus_unclassified	**35.98**	5.30	1.1E-11	1.9E-10
Lachnospiraceae_bacterium_5_1_57FAA	**41.44**	5.30	5.4E-15	2.0E-13
IBD T3 & IBD T2				
Streptococcus_gordonii	**-34.38**	5.74	2.1E-09	4.8E-08
Clostridium_leptum	**-27.02**	4.39	7.2E-10	2.1E-08
Alistipes_onderdonkii	**19.28**	4.64	3.3E-05	4.7E-04
Alistipes_finegoldii	**23.42**	4.14	1.5E-08	2.8E-07
Roseburia_inulinivorans	**25.46**	3.41	8.0E-14	3.0E-12
Eubacterium_eligens	**26.76**	3.35	1.4E-15	1.6E-13
Alistipes_shahii	**26.78**	3.42	4.7E-15	2.7E-13
Odoribacter_splanchnicus	**29.18**	5.71	3.2E-07	5.2E-06

Significance is defined by P-adjusted ≤ 0.05. Red colour indicates species with reduced abundance while green represents species with increased abundance in that pairwise comparison.

Supplementary table 11 Comparison of significant pathways in IBD dataset

A) Mean, standard deviation, and Kruskal-Wallis significance

Significant pathways in IBD	Control T0	IBD T0	IBD T2	IBD T3	p-value	p-adjusted
ARGDEG-III-PWY: L-arginine degradation IV (arginine decarboxylase/agmatine deiminase pathway)	9.9E-05 ± 2.8E-05	4.1E-05 ± 3.3E-05	2.6E-05 ± 6.3E-06	3.1E-05 ± 1.2E-05	5.4E-04	**3.2E-02**
PWY-43: putrescine biosynthesis II	9.9E-05 ± 2.8E-05	4.1E-05 ± 3.3E-05	2.6E-05 ± 6.3E-06	3.1E-05 ± 1.2E-05	5.4E-04	**3.2E-02**
PWY-7226: guanosine deoxyribonucleotides de novo biosynthesis I	1.3E-04 ± 3.9E-05	3.5E-04 ± 1.0E-04	3.0E-04 ± 4.1E-05	2.7E-04 ± 5.1E-05	3.6E-04	**3.2E-02**
PWY-7227: adenosine deoxyribonucleotides de novo biosynthesis	1.3E-04 ± 3.9E-05	3.5E-04 ± 1.0E-04	3.0E-04 ± 4.1E-05	2.7E-04 ± 5.1E-05	3.6E-04	**3.2E-02**
PWY0-1299: arginine dependent acid resistance	1.1E-04 ± 3.1E-05	2.5E-05 ± 1.6E-05	2.1E-05 ± 5.0E-06	2.5E-05 ± 1.3E-05	2.8E-04	**3.2E-02**
ASPARAGINE-DEG1-PWY: L-asparagine degradation I	2.4E-04 ± 5.1E-05	3.5E-04 ± 5.0E-05	4.1E-04 ± 7.5E-05	3.3E-04 ± 5.5E-05	7.2E-04	**3.6E-02**

Mean and Standard deviation (SD) is defined as mean ± SD. Significance is defined by P-adjusted ≤ 0.05

B) Significance of pairwise comparison from Dunn's test

Significant pathways in IBD	control_T0 vs IBD_T0	control_T0 vs IBD_T2	control_T0 vs IBD_T3	IBD_T0 vs IBD_T2	IBD_T0 vs IBD_T3	IBD_T2 vs IBD_T3
ARGDEG-III-PWY: L-arginine degradation IV (arginine decarboxylase/agmatine deiminase pathway)	**1.3E-02**	**3.3E-03**	**1.2E-02**	4.0E-01	4.3E-01	4.2E-01
PWY-43: putrescine biosynthesis II	**1.3E-02**	**3.3E-03**	**1.2E-02**	4.0E-01	4.3E-01	4.2E-01
PWY-7226: guanosine deoxyribonucleotides de novo biosynthesis I	2.9E-03	3.2E-03	3.2E-02	4.3E-01	3.9E-01	3.8E-01
PWY-7227: adenosine deoxyribonucleotides de novo biosynthesis	2.9E-03	3.2E-03	3.2E-02	4.3E-01	3.9E-01	3.8E-01
PWY0-1299: arginine dependent acid resistance	6.1E-03	4.0E-03	6.9E-03	5.9E-01	4.5E-01	5.4E-01
ASPARAGINE-DEG1-PWY: L-asparagine degradation I	1.2E-02	1.7E-03	6.0E-02	3.2E-01	3.5E-01	2.5E-01

Significance is defined by P-adjusted ≤ 0.05 for all pairwise comparisons

Chapter IV

IPCO: Inference of Pathways from Co-variance analysis

This chapter has been published as

Mrinmoy Das, Tarini Shankar Ghosh, Ian B. Jeffery

IPCO: Inference of Pathways from Co-variance analysis

February 2020, BMC Bioinformatics

https://doi.org/10.1186/s12859-020-3404-2

Contents

Chapter 4

Chapter 4 IPCO: Inference of Pathways from Co-variance analysis

4.1 Abstract

Background Key aspects of microbiome research are the accurate identification of taxa and the profiling of their functionality. Amplicon profiling based on the 16S ribosomal DNA sequence is a ubiquitous technique to identify and profile the abundance of the various taxa. However, it does not provide information on their encoded functionality. Predictive tools that can accurately extrapolate the functional information of a microbiome based on taxonomic profile composition are essential. At present, the applicability of these tools is limited due to requirement of reference genomes from known species. We present IPCO (Inference of Pathways from Co-variance analysis), a new method of inferring functionality for 16S-based microbiome profiles independent of reference genomes. IPCO utilises the biological co-variance observed between paired taxonomic and functional profiles and co-varies it with the queried dataset.

Results IPCO outperforms other established methods both in terms of sample and feature profile prediction. Validation results confirmed that IPCO can replicate observed biological associations between shotgun and metabolite profiles. Comparative analysis of predicted functionality profiles with other popular 16S-based functional prediction tools showed that the established tools had significantly lower performances with predicted functionality showing little to no correlation with paired shotgun features across samples.

Conclusions IPCO can infer functionality from 16S datasets and significantly outperforms existing tools. IPCO is implemented in R and available from **https://github.com/IPCO-Rlibrary/IPCO**.

Key words: microbiome, prediction, functionality, co-variance, novel method

4.2 Introduction

Microbiome research has expanded exponentially over the last decade and has shown that microbiota communities have significant roles in health maintenance, as well as being key inputs into food and industrial processes (Hadrich, 2018;Team, 2019). The study of microbiome communities fundamentally falls under two strategies: the taxonomic composition is determined by either amplicon sequencing (16S marker gene) or metagenomic whole genome shotgun sequencing (mWGS) with the latter providing additional information on the functional capabilities which allows the identification of genes and pathways. Despite the availability of mWGS, amplicon sequencing still remains popular due to its relatively low cost, quicker computation time, lower disk space requirements, and ability to detect a diverse set of taxa, including those with a low abundance based only on the marker gene. Comparisons of these two approaches have discussed both the advantages and disadvantages of these methods (Clooney et al., 2016;Jovel et al., 2016).

Amplicon sequencing is limited to providing only taxonomic information of the microbial communities. A number of tools can be used to predict the functional potential of the microbial communities obtained from 16S sequencing, the most cited being PICRUSt which was published as early as 2013 (Langille et al., 2013). Other widely used tools that were developed later are Tax4Fun (2015) (Asshauer et al., 2015), and Piphillin (2016) (Iwai et al., 2016). All of these tools rely on functionally annotated reference genomes. The difference between them is the methodology used to map the amplicon data to these references and the approach used to assign functional annotation when suitable reference genomes are not available. PICRUSt works by considering the phylogenetic tree and the distance to the closest functionally annotated reference microbe (Langille et al., 2013). It relies on the GreenGenes

database (DeSantis et al., 2006) for matching the references to the queried amplicon data. The major limitation of PICRUSt comes forth in the case of 16S sequences, which do not have sequenced/annotated genomes of phylogenetically close relatives in the reference database. Tax4Fun and Piphillin implement BLAST and global alignment respectively between the amplicon data and the reference genomes obtained from different databases (Asshauer et al., 2015;Iwai et al., 2016). All of these tools use KEGG orthologs (KOs) (Kanehisa and Goto, 2000) annotation from reference genomes which is combined with amplicon abundance to predict the functionality. The limitation of all these methods is the requirement of sequenced/annotated reference genomes.

To overcome constraints of limited reference taxa, an alternative approach would be to use pairs of 16S amplicon and metagenomic datasets which show co-variance between the samples based on taxonomic abundance and functionality respectively. The identified co-variance trends can then be combined with the taxonomic abundance of the queried dataset for whom the functionality will be predicted. In this paper, we present IPCO, a tool based on this novel approach for inferring functionality of a 16S amplicon dataset. The primary advantage of this new method is that it does not depend on the presence of sequenced and annotated genomes directly. IPCO is an application where values are assigned as the functional profiles for the samples of a 16S amplicon dataset based in a double co-inertia analysis involving the RLQ method (R-mode; Q-mode; and L-link between R and Q) (Dolédec et al., 1996;Dray and Legendre, 2008) between a paired taxonomic and functional dataset and a queried 16S amplicon dataset for which functionality will be inferred. Co-inertia analysis measures the concordance between two datasets, and maximises the squared covariance projected by two datasets (Dolédec and Chessel, 1994;Dray et

al., 2003). In paired taxonomic and functional profile datasets one would expect that alterations in the taxonomic profiles naturally should also reflect changes in its functional potential. Co-inertia can be further extended by application of the RLQ method, which integrates a third dataset (amplicon dataset in this case) and therefore analyses the co-inertia of the three datasets simultaneously. This methodology can provide a set of scores for the functional dataset and the amplicon dataset weighted by the paired taxonomic dataset.

IPCO's performance is compared with PICRUSt, Tax4Fun and Piphillin in terms of both sample and feature correlation with KEGG pathways from experimental datasets. IPCO also predicts MetaCyc pathway profiles and these predictions are validated against a paired mWGS dataset. Correlation of mWGS functional profiles against paired bile acids and short chain fatty acids (SCFAs) metabolite profiles confirmed the metabolomic associations with the observed metagenomic pathways. IPCOs ability to reproduce these biological associations is validated against these observed biological associations.

4.3 Methods

4.3.1 IPCO algorithm

IPCO is an implementation of the RLQ analysis which is also known as fourth corner analysis. It requires a reference taxonomic and functional paired dataset along with a third dataset, which is the 16S dataset for which the functional potential will be inferred. RLQ analysis is a double co-inertia method which explores two datasets (R and Q) through a mediator dataset (L). IPCO implements RLQ to associate the functional profiles (R) with a 16S profile dataset (Q) which is the 16S dataset for which functions need to be inferred through a mediator taxonomic profile dataset (L). R and L datasets are related as they have the same samples (paired) and are used as reference datasets. Q and L datasets are related as they have the same taxa identifiers. Functional profiles of R and taxonomic profiles of Q are standardised and scaled through the weighted average where the weights of the samples and taxa are obtained from L dataset. Through RLQ methodology, we obtain a R'LQ product table, which an association matrix of R and Q mediated through L abundance. In IPCO, we re-standardise the R'LQ products by adding the weighted average of the functional potential back to the association matrix to obtain inferred functional profiles for the samples of Q dataset.

In summary, IPCO implements the following steps:

L = Matrix from correspondence analysis of reference taxa table (L table)

R = Matrix from PCA of reference functional table (R table) weighted by samples from L

Q = Matrix from PCA of query taxa table (Q table) weighted by taxa from L

RLQ product = RLQ analysis (R, L, Q) (as described in the original paper)

rw = row weights from correspondence analysis of L

waR = Weighted average of R table given by (∑R[, j] * rw) / ∑rw (note: removed from RLQ product in RLQ calculation)

where j represents 1 to nth sample

Inferred profiles = RLQ product [i,] + waR

where i represents 1 to nth feature

4.3.2 Data collection

In the current study, human microbiome taxonomic and functional profile datasets were obtained from the HMP project (Human Microbiome Project, 2012a;b) using the curatedMetagenomicData R library (v.1.10.0) (Pasolli et al., 2017). mWGS functional (UniRef genefamilies) and taxonomic profiles datasets were obtained using the R library curatedMetagenomicData and paired V3-V5 16S rRNA OTU table was obtained from 16SHMPData R library (v.1.2.1) (Schiffer et al., 2019). Paired datasets were obtained for nasal, oral (buccal cavity), skin and stool samples. Representative OTUs of V3-V5 regions were downloaded from the HMP website (Human Microbiome Project, 2012a;b).

A larger reference dataset consisting of functional and taxonomic profiles generated from only mWGS data were also obtained from the curatadMetagenomicData. This set is comprised of 1180 healthy samples from various cohorts as described in Ghosh *et al.* (Ghosh et al., 2020).

Paired 16S and mWGS of an environmental dataset (Brazilian river water) used in this study are described in Tessler *et al.* (Tessler et al., 2017) and downloaded from the NCBI SRA (PRJNA389803, PRJNA310230). This cohort comprised of paired 16S rRNA and mWGS data obtained from four major rivers in Brazil: Amazon, Araguaia, Paraná, and Pantanal. The 16S rRNA and mWGS sequences was quality

filtered using Trimmomatic (v.0.38) (Bolger et al., 2014). Using USEARCH (v8.1), the quality filtered 16S rRNA sequences were dereplicated, clustered at 97% identity and chimera filtered (*de-novo* and using ChimeraSlayer) to obtain representative OTU sequences. Quality filtered reads were mapped to these OTUs to obtain the OTU table. The quality filtered mWGS data was processed using HUMAnN2 (v. 0.7.1) (Franzosa et al., 2018) to obtain mWGS derived taxonomic and functional profiles.

MetaCyc and KEGG pathway mapping files as provided with HUMAnN2 and HUMAnN1 were filtered to remove all known eukaryotic pathways. All samples from all datasets (UniRef gene profiles) were processed against the filtered MetaCyc mapping file to obtain the MetaCyc pathway abundance and coverage datasets.

The UniRef genefamilies dataset for all the samples were regrouped to KEGG Orthologs (KOs) IDs using humann2_regroup_table.py script and the KEGG to UniRef mapping provided in HUMAnN2 utilities. The regrouped KOs were processed using HUMAnN2 using the filtered HUMAnN1 KEGG pathways legacy database to obtain KEGG pathway profiles for the mWGS dataset.

All OTU datasets were filtered to remove samples with a sequencing depth of less than 1000 reads. Samples removed from OTU datasets were omitted from their paired mWGS datasets also. Normalised unstratified functional information was used in the implementation after the removal of UNMAPPED, UNGROUPED and UNINTEGRATED variables.

4.3.3 Data normalisation and transformations

Taxonomic and functional abundance datasets were transformed using the following transformations: Z-scaling, proportion normalisation, log10 on rarefied and log10 on proportional data with $1e10^{-5}$ added as minimum count value, Hellinger

transformation (Legendre and Gallagher, 2001) and centred log ratio (clr) transformation (Aitchison, 1982). These transformations were investigated to identify the transformation best suited for IPCO.

4.3.4 Validation of IPCO predictions

IPCO is dependent on reference paired functional and taxonomic datasets. To validate the methodology, a bootstrap strategy was implemented to evaluate its predictions (**Figure 1**). A subset of the samples from the 16S table were randomly selected and considered as table Q. The samples omitted in Q formed the taxonomic table L and were matched with its pathway abundance dataset to obtain a paired functional (R) and taxonomy (L) datasets thus removing pathway information for the samples present in table Q. Using IPCO on R, L and Q table, pathway profiles were obtained for the samples from Q. Both inferred sample and feature (pathways) values were correlated using Spearman correlation with the actual mWGS pathway dataset for those samples present in Q. The bootstrapping analysis was repeated with 100 iterations to randomly subsample the reference datasets. An average was taken for both inferred sample to actual sample and inferred pathway to actual pathway correlation values from the 100 iterations.

To ensure that the predictions are not due to presences of homogenous samples and/or overfitting, the predicted profiles at each bootstrap were compared against shuffled mWGS sample profiles also. This was carried out using the HMP stool KEGG and MetaCyc pathways and its paired OTU level taxonomic dataset at all reference/test split described above.

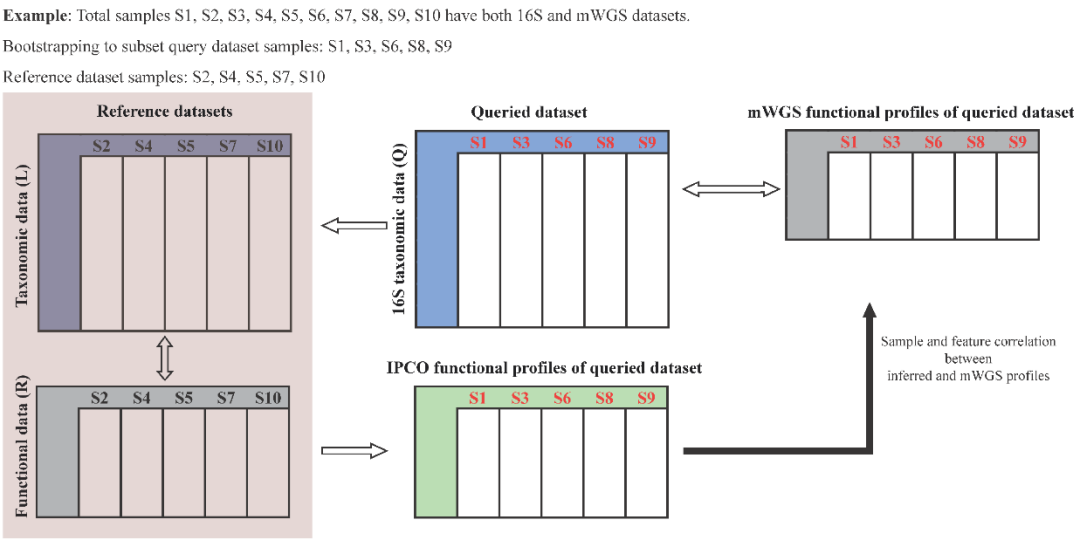


Figure 1 Implementation of IPCO and bootstrap iterations
The datasets were randomly subsampled into query and reference dataset. Reference dataset consist of taxonomic and functional profiles for the same samples. Reference taxonomic and queried dataset consist of different samples but are mapped to same taxa. The inferred profiles are correlated against the mWGS functional profiles for the query dataset to obtain the degree of associations. 100 bootstrap iteration were carried out to randomly generate different subsamples and measure the degree of association for each iteration.

4.3.5 Effect of reference dataset size and taxonomic dataset

IPCO is based on covariance between the datasets. The size of the reference dataset and the taxonomy used will both affect the covariance between the taxonomic and functional datasets. IPCO was implemented at various taxonomic levels and various reference dataset sizes by subsampling the reference (10%, 30%, 50%, 70% and 80%).

Taxonomy was assigned to the representative sequences using the RDP database (v.11.4) (Cole et al., 2014) implemented in mothur (v.1.34.4) (Schloss et al., 2009). In addition, SPINGO (Allard et al., 2015) was used for species assignment using the RDP database (v.11.2). All levels of taxonomic classification were classified to a threshold of ≥ 80% confidence. At any level, if the classified threshold was below 80%, it was set as unclassified.

To ensure that the predictions are not due to the presence of homogenous samples and/or overfitting, the predicted profiles at each bootstrap were compared

against shuffled wMGS sample profiles also (Labels on the R and L reference datasets were shuffled). This was carried out using the HMP stool KEGG and MetaCyc pathways and its paired OTU level taxonomic dataset at all reference/test split described above.

4.3.6 Functional prediction with published methodology

For PICRUSt, the representative 16S rRNA sequences (OTUs) were mapped to Greengenes v13.5 to obtain closed representative OTU tables and processed through the PICRUSt pipeline to obtain KO profiles.

For Tax4Fun, the taxonomic information for the OTUs were obtained from the SILVA123 reference downloaded from Tax4Fun website. The OTU table with taxonomic profiles were processed through Tax4Fun R library to obtain KO profiles. All processing was carried out with default settings.

For Piphillin, OTU dataset and the representative OTU sequences were formatted as per requirement for Piphillin and all files were uploaded to the Piphillin website. An identity threshold of 90% was used and the KEGG database October 2018 version was used. The threshold of 90% was used as the default of 97% returned insufficient hits to reference genomes.

The KEGG pathway abundances were obtained using HUMAnN2 by processing the predicted KO profiles using the filtered HUMAnN1 KEGG legacy database. This allowed consistency in pathway calculation across different methods.

Sample-to-sample and feature-to-feature (KOs and KEGG pathways) abundance correlations were calculated by comparing the predicted values obtained from the various tools against the mWGS generated profiles using Spearman

correlation. KOs and KEGG pathways present in both predicted and its paired mWGS datasets were considered.

4.3.7 Comparison between IPCO at KO and KEGG level with published tools

The Sample-to-sample and feature-to-feature correlations observed using the published tools for all the sites were compared against the correlation observed for IPCO inferred KO and KEGG pathways using reference:test split of 50:50 at all sites. Further, the IPCO inferred KO profiles (faecal site, reference:test 50:50) at each bootstrap were processed using HUMAnN2 to obtain inferred-KO KEGG pathways which were correlated against its paired mWGS KEGG pathway profiles. This was done to investigate the accuracy of using the IPCO inferred KO profiles to determine functional abundance at KEGG pathway level.

4.3.8 Comparison of inference capability from mWGS species and 16S species level as reference taxonomy on an external query dataset

It was investigated whether using mWGS species level dataset is sufficient as a reference set (table L) to generate inferred functionality for the independent dataset. The independent dataset used was the ELDERMET community dataset (Claesson et al., 2012). If this generated comparable results as 16S taxonomic reference, then it would allow generation of a larger reference R and L table by adding more samples from curatedMetagenomicData hub which would incorporate more functionality else would have to use HMP or similar 16S dataset as table L limiting the reference dataset size. Species level dataset were obtained for the 16S ELDERMET and HMP 16S as described earlier. IPCO was implemented with the R table as HMP pathways dataset, L table as HMP closed OTU or 16S species or HMP mWGS species dataset and Q

table as ELDERMET 16S closed OTU or species dataset. Samples and common features from inferred table were correlated with the paired elder mWGS functional dataset.

The species and functional dataset for the 1180 healthy samples were obtained from curatedMetagenomicData. The selection of healthy samples is described in Ghosh *et al.* (Ghosh et al., 2020). This dataset is referred as reference healthy in the results. IPCO was implemented with R table as reference pathways, L table as species mWGS dataset and Q table as ELDERMET species 16S dataset. The inferred functionality dataset for the ELDERMET dataset was correlated with mWGS functionality dataset to obtain sample and feature correlations. Correlations obtained from using this healthy reference were compared with the correlation values from using HMP as reference.

4.3.9 Validation with independent datasets

Validations of the IPCO's prediction was carried out with the following assumptions: The healthy functional and taxonomic references from a particular site can be used to predict diseased samples also obtained from the same site without requiring diseased sample profiles as reference. This would confirm the appropriateness of using provided references. The second assumption is that the biological agreement between measured metabolite levels and mWGS functional profiles is reflected in the inferred functional profiles. To validate the results on external independent datasets, analyses were carried out on two cohorts: a colorectal cancer (CRC) publicly available data (Zeller et al., 2014) and an in-house elderly community data (Claesson et al., 2012). The CRC dataset was used to validate the first assumption. The CRC cohort contains paired 16S and mWGS sequences obtained

from faecal samples of healthy and CRC individuals. The mWGS taxonomic and functional profiles were obtained from curatedMetagenomicData for the healthy (n=50) and CRC samples (n=41) based on the disease stratification provided. The forward reads from the 16S dataset were quality filtered using Trimmomatic. The quality filtered reads were processed using USEARCH and taxonomically annotated as described earlier. Predicted functional profiles (KEGG pathways) of the 16S CRC samples were also obtained for PICRUSt, Tax4Fun and Piphillin as described earlier.

First, we assessed whether IPCO can be applied to infer the functional potential from the 16S microbiota profiles of diseased samples or not using the CRC cohort. The 16S genus level and mWGS pathways (KEGG and MetaCyc) profiles of the CRC samples were used as reference in IPCO for inferring functionality of the queried 16S CRC samples. This was carried out using the bootstrap approach (100 iterations) described earlier to ensure the reference and test datasets (reference/test split 50:50) did not contain the same CRC samples. This would serve as the baseline for all other comparisons where the applicability of the healthy samples as reference would be determined. Next, using the taxonomic and paired functional datasets of the healthy samples from the same cohort as reference for IPCO, functionalities were inferred for the 16S genus profiles of the CRC samples. The genus level reference dataset was used as a mediator as most 16S datasets are usually classified down to genus level. Both 16S and mWGS genus dataset from the in-cohort healthy were used as reference. Finally, the external healthy taxonomic and functional profiles provided in IPCO were used as reference to infer the functionality of the 16S CRC samples. The healthy samples from the CRC samples were also present in the IPCO healthy reference and were removed before implementing IPCO. Using multiple reference datasets with IPCO methodology allowed comparing the differences in predictive capacity observed

with using different reference datasets (that includes in-cohort and external data). Further, functionalities predicted for 16S CRC samples using PICRUSt, Tax4Fun and Piphillin were also included to compare the differences in prediction observed in different tools. Correlation observed (Sample-to-sample and feature-to-feature) using the CRC samples as reference vs healthy samples as reference were compared for IPCO to determine the appropriateness of IPCO's reference. Correlation observed (Sample-to-sample and feature-to-feature) between predicted functional profiles obtained using 16S CRC samples and mWGS CRC samples using published tools were also compared with IPCO to evaluate the performance of all the tools.

The elderly cohort contains mWGS, 16S, and metabolome datasets for the same samples. This was used to validate the second assumption: biological agreement between metabolite profile and mWGS functional potential (KEGG and MetaCyc pathways) is replicated in inferred functional profiles also. To investigate the biological signal in the predicted functionalities, functional profiles were obtained from all tools.

For IPCO, paired functionality (KEGG pathway and MetaCyc) and taxonomy at species level from the reference healthy datasets provided in IPCO was used to infer functionalities for the 16S elderly samples. Functional profiles (KEGG pathways) were obtained using PICRUSt, Tax4Fun and Piphillin using the same approach as described earlier for the elderly community 16S dataset. Two types of metabolites: bile acids and short chain fatty acids (SCFAs) which are widely studied in microbiome research were considered for investigation. The metabolite dataset was log10 transformed on the measured metabolite level after adding 1e-05 as minimum count value. The mWGS functional profiles were correlated with the metabolite profiles and the directionality, degree of association and significance was noted. Significance of

mWGS profile correlation was determined by p-adjusted ≤ 0.05 unless stated otherwise. The inferred functions obtained from all the tools including IPCO were then correlated with the same metabolites and the results were compared with mWGS results to investigate the direction, correlation, and significance (p-value ≤ 0.1). Only key pathways responsible for these metabolites were considered and agreement with mWGS results in terms of directionality and significance were considered to be correct, with a change in directionality or non-significance being considered as false positives.

4.3.10 Statistical analysis

All analysis was carried out in R (v.3.5.1) (Team, 2018). All correlations measured were carried out using Spearman correlation. Kruskal-Wallis test was used as applicable. Dunn's test using dunn.test library (v.1.3.5) (Dinno, 2017) was used for pairwise comparison at different taxa and sample threshold levels. P-value adjustment was carried out using Benjamini-Hochberg procedure. Covariance between paired taxonomic and functional dataset was investigated with co-inertia analysis using ade4 (v.1.7.13) library (Dray and Dufour, 2007). Significance of co-inertia was determined with the ade4 randtest function. Plots were created using ggplot2 (v.3.1.0) (Wickham, 2009), RColorBrewer (v.1.1.2) (Neuwirth, 2014) and gridExtra (v.2.2.1) (Auguie, 2016) R libraries.

4.4 Results

4.4.1 Description of the study cohort

T**able 1** describes the cohort retained after removing samples with low sequencing depths and stratified functional features for the initial analysis. The samples retained were investigated using the IPCO, PICRUSt (Langille et al., 2013), Piphillin (Iwai et al., 2016) and Tax4Fun (Asshauer et al., 2015) to evaluate the performance of the tools.

Table 1 Number of mWGS features and samples retained in the initial analysis

	Samples	**KOs**	**KEGG pathways**	**MetaCyc**	**OTUs**
HMP nasal	61	5971	129	659	7464
HMP oral	71	5829	133	625	13696
HMP skin	8	4886	132	593	3492
HMP stool	87	6356	128	712	9966
Water (Brazilian rivers)	37	2724	114	497	1185

Total number of samples and features across the five reference datasets.

IPCO was initially implemented in the HMP (Human Microbiome Project, 2012a) stool cohort to investigate the effects of different data transformation, taxonomic levels and sample size thresholds.

4.4.2 Data transformation of the reference datasets affects IPCO results

In IPCO, the covariance between the three datasets (R, L and Q) vary depending on the transformation and normalisation method used. Investigation with the transformation/normalisation mentioned in the methods is shown in **supplementary figure 1** and **supplementary table 1**. The Hellinger transformation was best suited as it had a higher RV coefficient for both samples and features and similar correlation values compared to other methods. Observed RV coefficients were significant (p-

value ≤ 0.05) for all cases. Hence, this transformation is implemented as default for all further analysis.

4.4.3 Lowest taxonomic levels and high samples size provides best results with IPCO

Implementation of IPCO with 100 bootstraps for each subsampling at different reference sample sizes on both KEGG and MetaCyc pathway abundance datasets showed that the best sample and feature correlations were observed with the lowest taxonomic levels and highest sample size (**Figure 2A-B, supplementary figure 2A-B**). No significant difference was observed in the sample correlation for any reference dataset size except for between 10% and other reference sizes at family level in the MetaCyc dataset **(supplementary table 2)**. However, the feature correlation increased with increased sample size and at the lowest taxonomy levels. No significant differences were observed for feature correlation using a reference size of at least 30% or larger in the KEGG pathway analysis and 50% or more for MetaCyc at the different taxonomic levels investigated **(supplementary table 2)**.

The outlier observed in figure 1 and supplementary figure is due to increased abundances of pathways in that sample which are observed to be decreased in other samples. Looking at the KEGG pathways for the outlier samples, it was observed that the most abundant pathways also had a low coverage (maximum coverage = 1e-04).

The randomisation of mWGS samples through shuffling during evaluation at each iteration resulted in a complete lack of feature-to-feature correlation (**supplementary figure 3**). Although a high sample-to-sample correlation was observed due to the functional redundancy across samples, the pattern associated with

different thresholds and with the unshuffled predictions seen in **figure 1** and **supplementary figure 2** were lost.

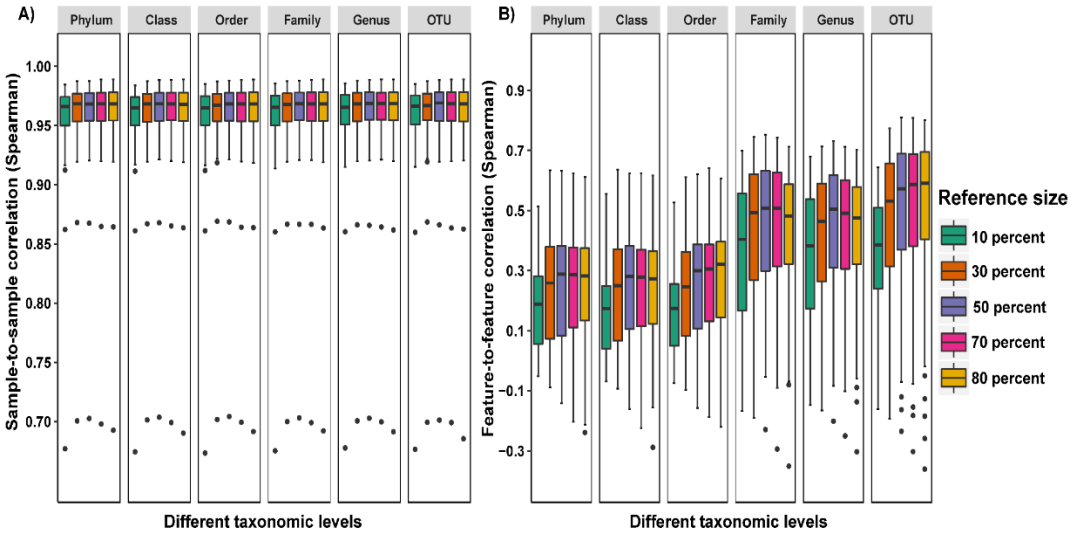


Figure 2 Sample and feature correlation using KEGG pathway abundance at different taxonomic levels and reference dataset size
Boxplots showing the variation of **A)** Sample to sample correlations **B)** Feature to feature correlations obtained between the inferred KEGG pathway abundances and the mWGS functional profiles at all different reference sizes and taxonomic levels.

The shuffling of mWGS query samples during evaluation at each iteration resulted in completed lack of feature-to-feature correlation (**supplementary figure 3**). Although a high sample-to-sample correlation was observed, the pattern associated with different threshold and with the unshuffled predictions seen in **figure 2** and **supplementary figure 2** were lost. This suggests that the homogeneity between reference and query and overfitting in reference may not be associated with the reference threshold splits.

4.4.4 IPCO outperforms PICRUSt, Tax4Fun and Piphillin in terms of both sample and feature correlations

To evaluate the functional inference of IPCO, we applied PICRUSt, Tax4Fun and Piphillin to the same datasets (**Table 1**) and KEGG pathway profiles were inferred. Spearman correlation was calculated for the inferred pathway profiles against its mWGS abundance both in terms of sample and feature correlation. IPCO outperformed PICRUSt, Tax4Fun and Piphillin in terms of sample correlation across all datasets (**Figure 3A, supplementary table 3**). IPCO showed highest sample correlation with a narrow IQR range for stool and oral samples. Skin and nasal dataset showed lower sample correlations compared to stool and oral samples, however it was observed to be higher than what was observed using the other tools. The lowest sample correlation was observed using the Brazilian river water dataset, but it was also higher than other tools for that site (**supplementary table 3**).

Upon investigating the feature-to-feature correlations, it was observed that IPCO outperforms PICRUSt and Tax4Fun in stool and Brazilian river water datasets (**Figure 3B**). Nasal, oral and skin dataset revealed a lack of correlation using IPCO. It was noted that across all datasets, the median feature correlation for PICRUSt, Tax4Fun and Piphillin was close to zero.

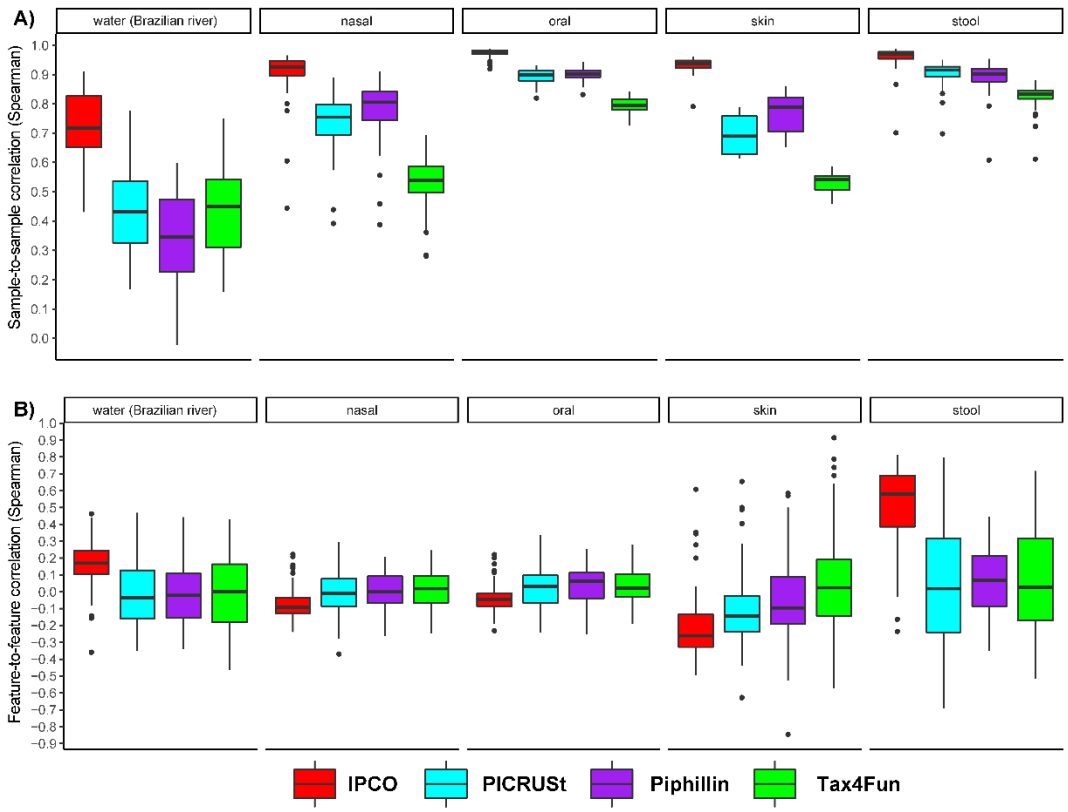


Figure 3 Sample and feature correlations between inferred KEGG pathways and mWGS KEGG pathways profiles at different sites and using different methods
Boxplots showing the comparison of **A)** Sample to sample correlations and **B)** Feature to feature correlations obtained between the inferred KEGG pathway abundance and the mWGS functional profiles at different sites using different methods.

Sample-to-sample and feature-to-feature correlations based on KO abundances obtained from IPCO, PICRUSt, Tax4Fun and Piphillin were also calculated. IPCO was observed to outperformed other methods in terms of correlation values between both inferred sample and feature profiles against the observed mWGS sample and feature profiles (**Figure 4, supplementary table 4**) for faecal and Brazilian river water datasets. Feature-to-feature correlations for the remaining sites (nasal, oral and skin) were poor with the median correlation being close to zero for all tools including IPCO.

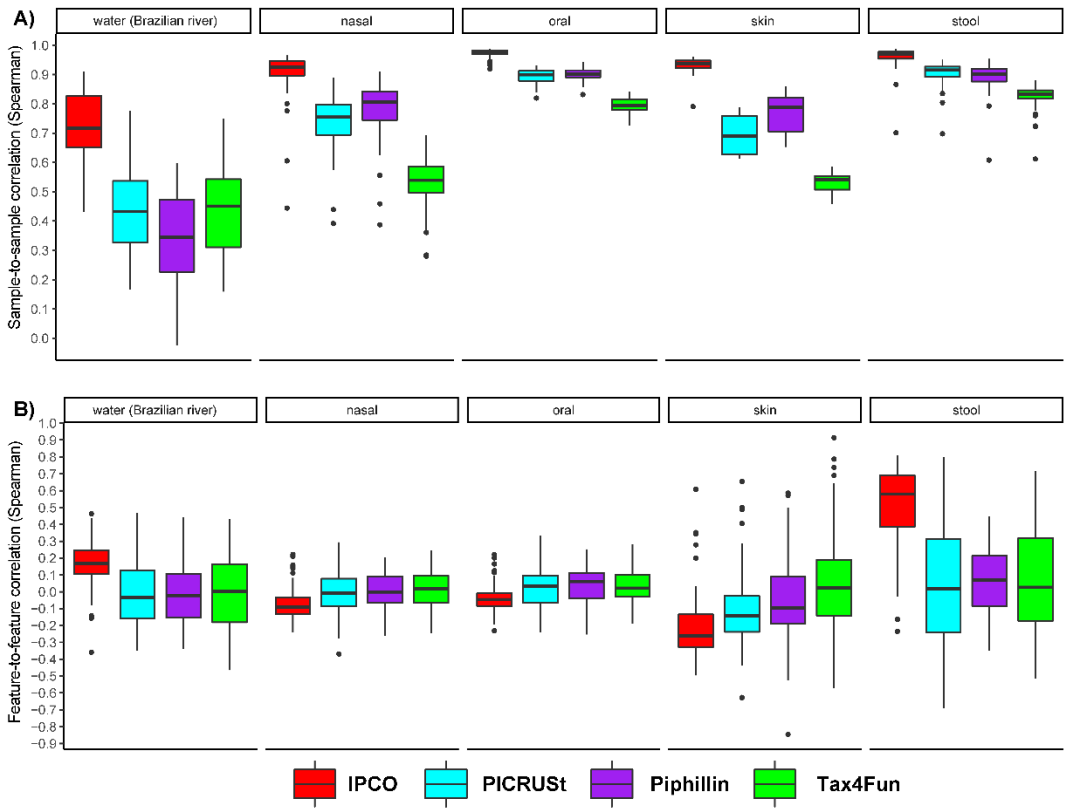


Figure 4 Sample and feature correlations between inferred KO and mWGS KO profiles at different sites and using different methods
Boxplots showing the comparison of **A)** Sample to sample correlations and **B)** Feature to feature correlations obtained between the inferred KOs and the mWGS KO profiles at different sites using different methods.

The IPCO inferred KO profiles obtained from HMP stool dataset using the equal reference to test split (50:50) threshold was further processed to obtain the inferred-KO KEGG pathway profiles at each iteration. Interestingly, these profiles at pathways levels showed a higher feature-to-feature correlation (**1st quartile:** 0.31, **median:** 0.49, **3rd quartile:** 0.67, **supplementary figure 4**) when compared to other published tools but lower than the default IPCO methodology as observed in figure 3.

4.4.5 Lack of significant covariance between taxonomy and functional datasets results in a lack of feature correlation in nasal, oral, skin, and Brazilian river water datasets

To investigate the poor performance by IPCO on the other sites excluding stool, the co-variance between the taxonomic and functional dataset was calculated. Co-inertia analysis of the taxonomic profiles with its paired functional datasets across all sites revealed a lack of significant covariance between taxonomy and functional profiles in the nasal, oral and skin datasets (**supplementary table 5**). The Brazilian river water dataset showed significant covariance when the mWGS derived taxonomy was used (while the 16S dataset did not co-vary significantly) (**supplementary table 5**). This may in part explain the poor performance of IPCO on these datasets. The taxonomic abundance of the reference dataset is not reflected in its paired functional dataset which resulted in a lack of significant covariance. It was observed that the functional diversity determined by the observed number of pathways had a narrow range compared the paired taxonomic diversity which showed more variation for all body sites excluding the stool dataset. This functional redundancy and the lack of detection of unique functionality suggests that the functional heterogeneity of species was not accurately reflected in these datasets. Given the lack of covariance in these datasets, further analysis was carried out using the stool dataset.

4.4.6 Higher pathway coverage improves functional inference

Investigation of the effect of pathway coverage on the correlation values between the observed pathway abundance and the inferred pathways obtained using IPCO showed that coverage correlated well with functional pathway prediction (**Figure 5A**). Based on this, the KEGG pathways were binned based on thresholds such that pathways

below the mean coverage of 0.01 were considered low correlation predictions and pathways with a mean coverage over 0.1% were considered high correlation predictions with the remaining pathways with a mean coverage between 0.01 and 0.1 being considered medium correlated predictions. These predictions showed correlation values between 0.25 - 0.6 whereas the high correlated predictions had correlation values between 0.6 - 0.7 (**Figure 5A)**. In case of MetaCyc pathways, we observed similar results where pathways with coverage less than 0.41 (1st quartile of mean coverage across samples) were low correlated predictions. Pathways whose mean coverage was between the 1st quartile (0.41) and the median value (0.99) showed improved feature correlation for the inferred pathways whereas the best feature correlation (0.37 - 0.62) was observed for those pathways whose average pathway coverage was greater than its median value (**Figure 5B)**. For both KEGG and MetaCyc, we were able to get high correlation values for more than 50% of the pathways based on the coverage filtering. The number of pathways binned into each of the coverage thresholds for both KEGG and MetaCyc are described in **table 2**. All reported observations were carried out on HMP stool functional dataset.

Table 2 Pathways retained at different coverage thresholds

KEGG coverage threshold	**Total**	**<0.01**	**0.01-0.1**	**>0.1**
Number of KEGG pathways	118	19	30	69
MetaCyc coverage threshold	**Total**	**<0.43**	**0.43-0.99**	**>0.99**
Number of MetaCyc pathways	693	173	166	354

Number of KEGG and MetaCyc pathways identified at different coverage thresholds. Coverage thresholds are strongly correlated to the accuracy of the inferred pathway profiles.

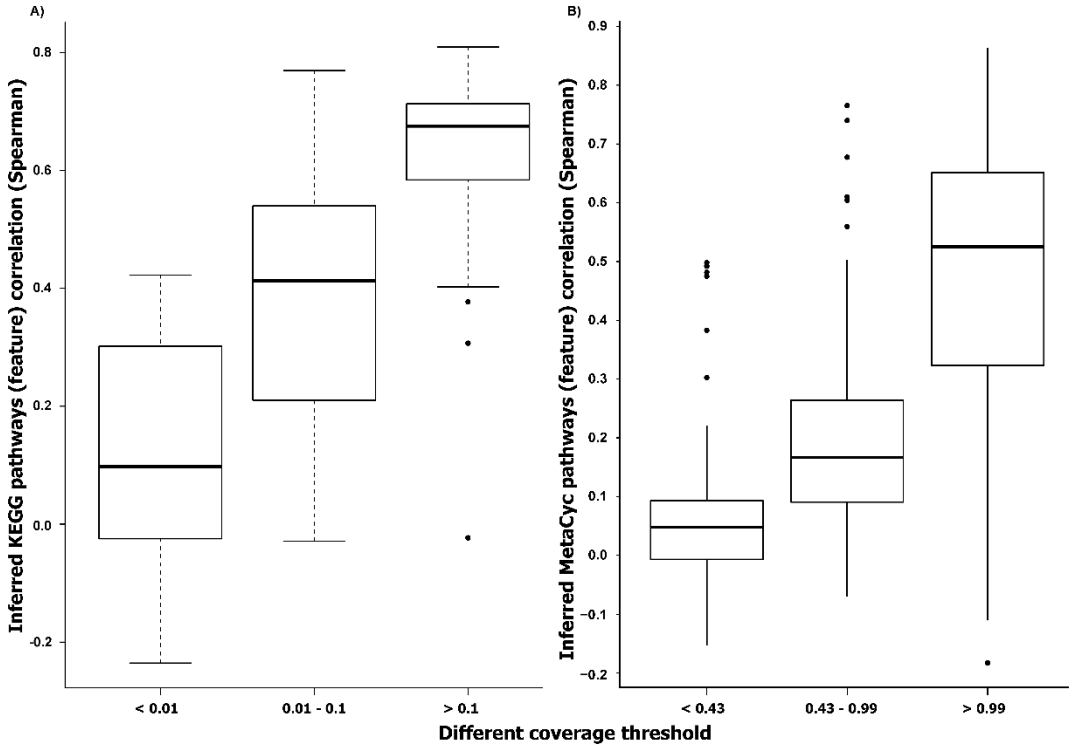


Figure 5 Comparison of feature-to-feature correlations at different pathway coverage thresholds for the KEGG and MetaCyc
Boxplot showing the functional feature-to-feature correlations obtained using different coverage thresholds for the **A)** KEGG pathways and **B)** MetaCyc pathways.

The mean coverage for the predicted KEGG pathways from PICRUSt, Tax4Fun and Piphillin modelled against its feature correlation values showed no positive association between coverage and feature correlation (**supplementary figure 5**). This indicates that the coverage filter was not applicable for these tools as opposed to IPCO. All subsequent analyses were carried out using all the functionalities without filtering for coverage.

4.4.7 IPCO can accurately infer sample and feature profiles using taxonomy from mWGS or 16S amplicon datasets

An alternative approach to using a reference 16S dataset and paired mWGS functional is to derive the taxonomic information from the reference mWGS dataset itself. Investigation of using taxonomic information derived from mWGS showed a

comparable performance to using a 16S species or closed OTU level dataset when inferring the functionality of an external 16S dataset (**supplementary figure 6**). **Table 3** describes the sample characteristics considered for this analysis. The reference dataset in this case consists of a large cohort of healthy samples as detailed in the methods (Ghosh et al., 2020). The validation dataset used is the ELDERMET dataset (Claesson et al., 2012).

Table 3 Number of samples, pathways, and species in the validation datasets

	samples	**KEGG pathways**	**MetaCyc**	**mWGS Species**	**16S Species**	**Closed OTU**
Validation	79	123	776	NA	201	842
Reference healthy	1180	143	833	772	NA	NA
HMP	87	128	712	353	282	1341

The number of samples and features present in the reference healthy and validation (ELDERMET) datasets, which were used when inferring functional profiles and comparing with observed functional profiles of the validation dataset. NA; Not applicable.

It was observed that the use of species levels datasets obtained from the same mWGS data that was used to compute metagenomic functional profiles was sufficient to infer functionality for 16S datasets (**supplementary figure 6**). Further validation involving the replication of biological pathway to metabolite associations was carried out using the reference healthy functional and paired mWGS species profiles as reference in IPCO to infer the functionality of the ELDERMET 16S dataset.

4.4.8 Healthy references implemented with IPCO shows better inferences for diseased samples

IPCO was implemented with both in-cohort and the external healthy references on the CRC samples (16S genus level profiles) from the Zeller *et al.* The predicted pathways from the CRC dataset highlighted that the healthy references can be used to predict the diseased samples as determined by the high sample and feature correlation (**Figure 6, supplementary figure 7, supplementary table 6**). The healthy IPCO reference and in-cohort healthy reference samples resulted in higher predicted feature correlation (KEGG predicted CRC vs mWGS CRC) compared to using CRC dataset as reference itself. Repeating the analysis using MetaCyc, the healthy reference (in-cohort 16S genus profiles) showed higher feature correlation compared to using CRC as reference but there was no significant differences in the sample profile predictions.

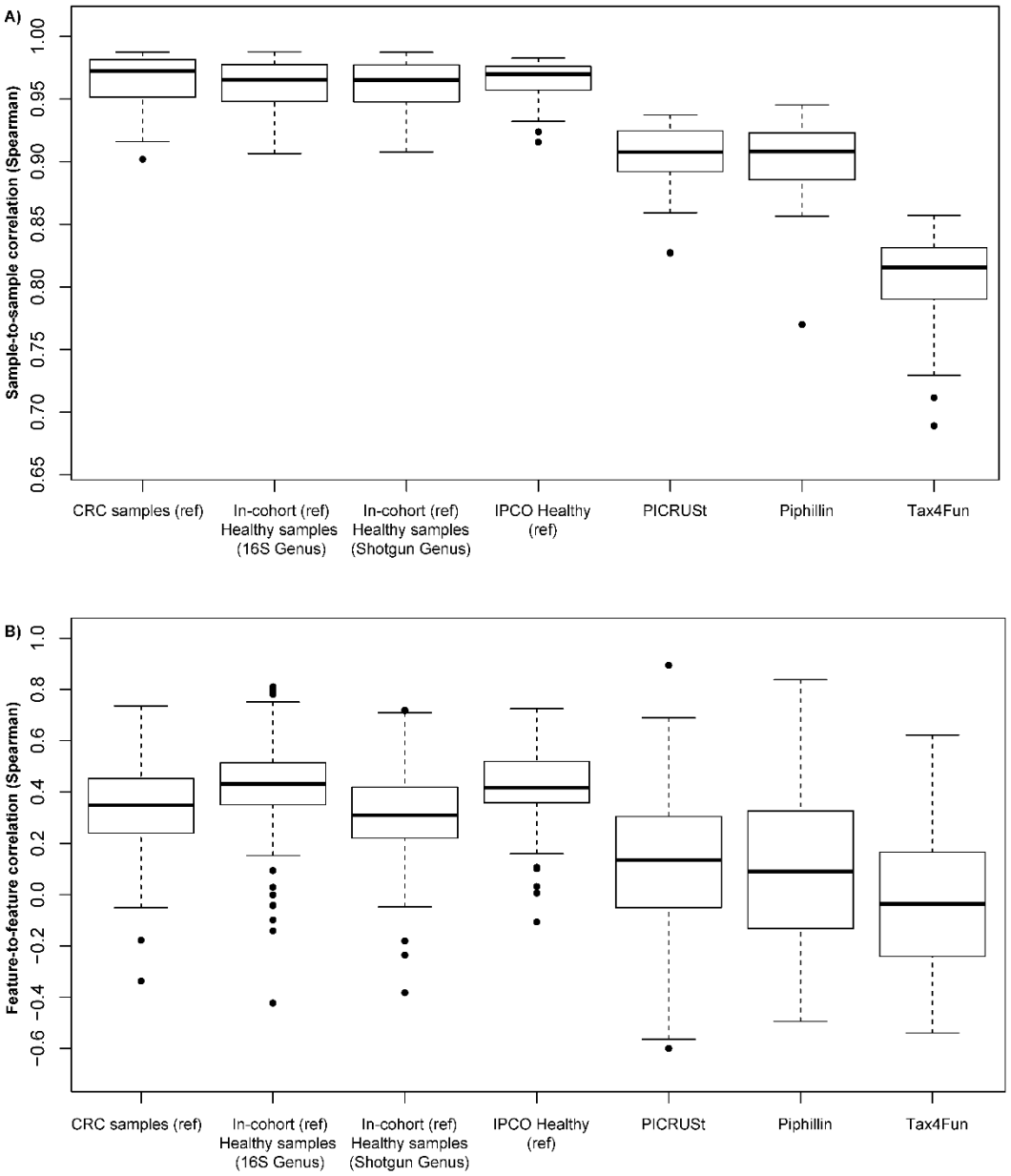


Figure 6 Comparison of sample and feature correlation observed when using different reference datasets to predict CRC sample functionality (KEGG pathways)

Boxplot showing **A)** Sample-to-sample correlation observed with the use of only CRC samples as reference, healthy samples from same cohort as reference where the reference taxonomic dataset is 16S genus profiles, healthy samples from the same cohort with mWGS genus profiles as reference taxonomy, IPCO reference healthy at genus level (taxonomy reference), PICRUSt, Piphillin and Tax4Fun. **B)** represents feature-to-feature correlation observed with the use of different references as described for **figure 6A**.

4.4.9 Inferred profiles from IPCO replicate mWGS functional pathway to metabolite profile associations

Correlation of the mWGS derived functionality (KEGG and MetaCyc) from ELDERMET samples to paired bile acid and SCFA profiles identified associations between key pathways and biologically relevant metabolites.

Investigation of the bile acid profiles showed that KEGG pathway "ko00121: Secondary bile acid biosynthesis" pathways significantly negatively correlated with primary bile acids (cholic acid and chenodeoxycholic acid) while "ko00790: Folate biosynthesis" known to promote bile acid levels (Delgado-Villa et al., 2009) is observed to be significantly positively correlated with primary bile acids in ELDERMET mWGS data.

With the secondary bile acids (lithocholic acid, dehydrocholic acid, 12-ketolithocholic acid, dehydrolithocholic acid, hyodeoxycholic acid and isolithocholic acid), it was observed that "ko00121: Secondary bile acid biosynthesis", "ko00430: Taurine and hypotaurine metabolism", "ko03070: Bacterial secretion system", "ko05100: Bacterial invasion of epithelial cells" were all significantly positively correlated with secondary bile acid levels. This validates the concordance between the measured bile acid profiles and the ELDERMENT mWGS functional profiles thereby with the functional activity of the microbiota. This concordance between the measured bile acid profiles and the ELDERMET mWGS functional profiles thereby validates the link between the functional activity of the microbiota and the biochemical characteristics of the gut environment.

The results of the inferred functional profiles obtained from all the tools showed that IPCO provides the best estimation of the observed associations between the mWGS dataset and bile acid profiles (**Figure 7, supplementary table 7**). It was

observed that for primary bile acids only PICRUSt showed significant correlation with "ko00121: Secondary bile acid biosynthesis", however the directionality was reversed. Neither Tax4Fun nor Piphillin showed significant associations for "ko00121: Secondary bile acid biosynthesis" abundance and primary bile acids. Looking at the secondary bile acids, we observed that 12-ketolithocholic acid did not show any significance with the inferred profiles obtained from all tools. Lithocholic acid and KEGG pathways obtained from IPCO agreed with mWGS results. Tax4Fun was significant but showed the opposite directionality to the observed association. Dehydrolithocholic acid was significantly associated with KEGG pathways in PICRUSt and Piphillin. All associations, directionality and significance from all tools compared to mWGS results are highlighted in **figure 7**. Overall, while correlating measured bile acids levels to predicted KEGG pathway profiles, IPCO was successful 62% of time, whereas PICRUSt, Tax4Fun and Piphillin were correct only 12%, 31% and 38% of time respectively.

Similar results were observed when the MetaCyc pathway abundance dataset was used as reference. Correlation of the ELDERMET mWGS MetaCyc pathways with bile acid profiles show significant correlation with "PWY-6518: glycocholate metabolism (bacteria)" and "1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis". These results were replicated with the inferred MetaCyc pathway profiles obtained from IPCO (**supplementary table 8**).

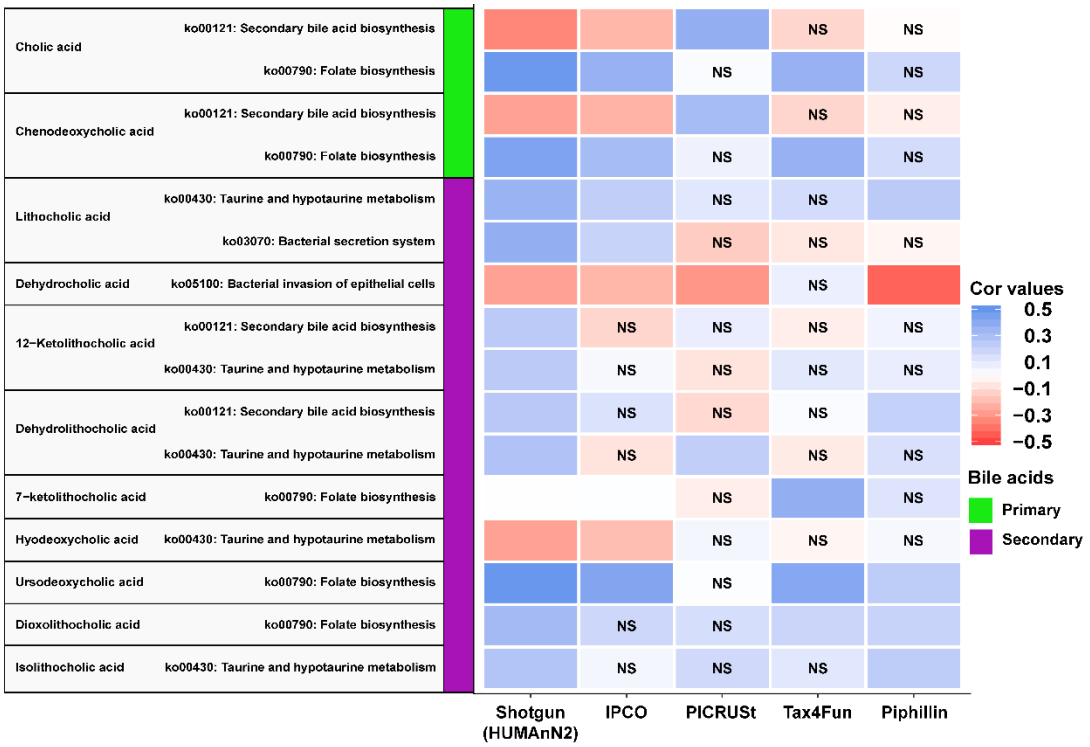


Figure 7 Correlation of the bile acids metabolite profiles with inferred KEGG pathway abundances from various methods

Correlation of the inferred bile acid metabolite profile to paired mWGS KEGG pathways shows significant association (p-adjusted ≤ 0.1) with known pathways as shown in 1st column. Directionality of association is shown by correlation values colour intensity. Pathways inferred from IPCO shows same directionality and significance (p-value ≤ 0.1) as observed with mWGS profiles for most cases. "**NS**" inside the cell represent non-significant (p-value > 0.1) associations.

The correlation between ELDERMET KEGG pathway abundance and SCFA (butyrate and propionate) profiles were observed to be not significant which included butanoate and propanoate metabolism, protein and amino acid metabolism (Lysine, Glumatine) pathways. This lack of association is consistent with the literature (Sze et al., 2019) and was replicated in the inferred profiles obtained from IPCO and PICRUSt (**supplementary table 9**). However, Tax4Fun and Piphillin showed significant associations for the inferred KEGG pathways obtained using those two tools for both butyrate and propionate levels. These significant associations are considered false positives, as they were not observed with the mWGS data. Piphillin reported the

highest number of false positive pathways for butanoate. In the case of propionate, IPCO also predicted two false positive pathways.

Investigation of the SCFAs (butyrate and propionate) levels with ELDERMET mWGS MetaCyc pathways replicated the KEGG analysis and again showed that butyrate shows no significant association with key butyrate MetaCyc pathways after adjustment which is consistent with the literature (Sze et al., 2019). This observation was replicated in the IPCO inferred functional profiles (**supplementary table 10**). Propionate also showed no significant association with mWGS pathways and this observation was replicated with IPCO inferred pathways (**supplementary table 10**). As other tools do not report MetaCyc pathway profiles, this investigation could not be carried out with other tools.

4.5 Discussion

We have developed IPCO, a novel tool which predicts functionality for 16S amplicon datasets but is not dependent on the direct mapping of 16S sequences to known reference genomes. Instead, it utilises paired mWGS functional and taxonomic datasets as references that are built from annotated genomes but do not assume that the functional potential of the taxa is the same as the reference genome.

Alterations at taxonomic level will affect the overall functional potential at community level (Eng and Borenstein, 2018). The robustness of the association between functionality and taxonomy is dependent on both the abundance of the various taxa and the distribution of function across these taxa. Using this concept, IPCO is able to utilise the biologically and statistically significant covariance observed between the reference taxonomic and functional datasets and infer the functional capabilities of an external 16S amplicon dataset. The IPCO implementation is also unique in that it provides a distinction between high quality predictions and lower quality predictions based on pathway coverage for both KEGG and MetaCyc pathways.

IPCO is reliant on the availability of reference datasets from the environment being studied. This raises the question of the appropriateness of using a set of defined samples (e.g. healthy reference) to infer samples that are dissimilar in some aspect (e.g. diseased) even within the same environment. By using different types of reference samples, it is observed that the healthy samples are better at predicting diseased samples as long as they are obtained from the same environment. In fact, in this analysis, the functional inference capability of IPCO for both KEGG and MetaCyc was better with the use of healthy samples from the same site (**Figure 6, Supplementary figure 7**)..

One weakness of available functional prediction tools is that using a reference set of known functions from only a set of known taxa limits the prediction of functionality for the amplicon dataset. This limitation results in a lack of feature-to-feature correlation i.e. KEGG pathway abundance calculated from 16S datasets do not correlate well across samples that are obtained from similar environments when compared with a paired mWGS dataset for the same samples. This is of concern as this potentially creates false positive results and/or reversed directionality when investigating functional profiles inferred from 16S datasets. By studying the IPCO inferred functionality and associations with biologically relevant metabolites, we have shown that the inferred functional capabilities obtained using IPCO, mimics the results of the mWGS functional profiles and outperform other predictive tools.

The filtering criteria in IPCO allows the users to select a set of functional pathways with sufficient coverage to be inferred. This removes functional pathways which may have been spuriously assigned due to the presence of only a small subset of genes/reactions. It is noteworthy that despite using a uniform method to tabulate the pathway level information for all the tools, the pathway coverage information could not be used for filtering out low-coverage pathways for any tools except IPCO. The lack of association between feature coverage and correlation observed in other tools may be due to the assumption made when mapping to functionally annotated references by the published tools. The reproducibility of results observed from both KEGG and MetaCyc shows that our method is independent of the functional annotation and can be implemented with any taxonomic level information and the taxonomic assignment can be done with any reference database as long as the taxa are present in the reference which acts as a mediator to co-vary the functional profiles with the 16S dataset. This may allow alternative implementations of the IPCO

methodology such as extrapolating of functional information for a set of samples obtained from shallow sequencing by using a subset of samples with deeper sequencing depth.

Overall, IPCO had a superior performance compared to other tools using both KOs and KEGG pathways, even when inferred KO-KEGG pathways was used. The feature correlation observed across all tools including IPCO were lower using KEGG pathway level predictions. This could be potentially due to increased sparseness at KO level compared to pathway levels. The processing time also increases drastically due to the increased size of functional reference dataset (Number of KEGG pathways in HMP stool (n=87) dataset is 128 whereas the number of KOs for the same dataset is 6356) which would be a limitation when using on a personal laptops or system with low computational and memory capacity. Similarly, the use of UniRef gene profile datasets is not feasible currently as nearly 2 million UniRef genefamilies are detected for the HMP stool samples alone which would require long processing times would require a high performance computation system to process such a large dataset..

IPCO performs better than the other established tools but it is not without its limitations. IPCO is reliant on a paired mWGS functional and taxonomic reference datasets which relay on functionally annotated genomes. As with other tools, sample profile predictions will appear to be highly correlated to the actual sample profiles due to functional redundancy at the pathway level where highly abundant pathways are shared across multiple taxa. Therefore, care should be taken when interpreting the predicted sample profiles. IPCO assigns a small pseudo value to each functionality due to the way the R'LQ algorithm calculates double co-inertia, which makes the resulting inferred functionality a non-zero abundance. To overcome this limitation, the low abundant functionality can easily be filtered by removing those functions whose

average inferred abundance across samples is below a certain quantile determined by the user. Although other tools predicted the features poorly across all sites, IPCO also performed poorly on the non-gut samples. Our analysis showed that this was due to a lack of suitable reference datasets (lacking significant co-variance between reference taxa and functions), but IPCO can be easily tuned to work at other body sites or environments as suitable mWGS data from these different environments become available. This is noteworthy as inferring functionality using amplicon-based approaches rely on its concordance with the functional profile, which is not possible if the functional-level distribution is discordant with the taxonomy. Why this would be the case is beyond the scope of this manuscript. The use of IPCO is also limited to potentially only those environments, for which as a reliable reference (i.e. significant covariance between taxonomic and functional profiles) is available. Samples from environments (e.g. low biomass) that may lack mWGS dataset limits the use of IPCO.

4.6 Conclusion

IPCO provides a novel approach for functional inference, which is not directly dependent on the availability of functionally annotated reference genomes. The IPCO inferred functionality profiles reflect the true observed biological functionality. IPCO can be easily implemented with the default datasets or with in-house reference datasets without relying on the external reference datasets. Overall, IPCO provides a reliable inference of functional potential and can be easily implemented in the R statistical software.

4.7 Declaration

Ethics approval and consent to participate Not applicable

Consent for publication Not applicable

Availability of data and materials R library, data and documentation are available at https://github.com/IPCO-Rlibrary/IPCO.

All reference datasets are obtained from curatedMetagenomicData R library and HMP1 project

Project name: IPCO-Rlibrary

Operating system(s): Platform independent

Programming language: R

Other requirements: ade4 R library

License: ≥ GNU version 2

Restrictions for use by non-academics: None

Competing Interests None to declare

Funding IBJ and MD are funded by an SFI Starting Investigator Research Grant (SIRG) (13/SIRG/2128). TSG is supported by Science Foundation Ireland (APC/SFI/12/RC/2273) in the form of a research centre, APC Microbiome Ireland.

Contributors MD and TSG carried out the bioinformatics analysis, compilation, and interpretation of the work, drafting and revising of the manuscript submitted. IBJ contributed to the concept and design of the work, interpretation of the work, drafting and revising of the manuscript submitted.

4.8 References

1. Aitchison, J. (1982). The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society: Series B (Methodological)* 44**,** 139-160.
2. Allard, G., Ryan, F.J., Jeffery, I.B., and Claesson, M.J. (2015). SPINGO: a rapid species-classifier for microbial amplicon sequences. *BMC Bioinformatics* 16**,** 324.
3. Asshauer, K.P., Wemheuer, B., Daniel, R., and Meinicke, P. (2015). Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics* 31**,** 2882-2884.
4. Auguie, B. (2016). gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.2.1.
5. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30**,** 2114-2120.
6. Claesson, M.J., Jeffery, I.B., Conde, S., Power, S.E., O'connor, E.M., Cusack, S., Harris, H.M., Coakley, M., Lakshminarayanan, B., O'sullivan, O., Fitzgerald, G.F., Deane, J., O'connor, M., Harnedy, N., O'connor, K., O'mahony, D., Van Sinderen, D., Wallace, M., Brennan, L., Stanton, C., Marchesi, J.R., Fitzgerald, A.P., Shanahan, F., Hill, C., Ross, R.P., and O'toole, P.W. (2012). Gut microbiota composition correlates with diet and health in the elderly. *Nature* 488**,** 178-184.
7. Clooney, A.G., Fouhy, F., Sleator, R.D., A, O.D., Stanton, C., Cotter, P.D., and Claesson, M.J. (2016). Comparing Apples and Oranges?: Next Generation Sequencing and Its Impact on Microbiome Analysis. *PLoS One* 11**,** e0148028.
8. Cole, J.R., Wang, Q., Fish, J.A., Chai, B., Mcgarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., and Tiedje, J.M. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 42**,** D633-642.
9. Delgado-Villa, M.J., Ojeda, M.L., Rubio, J.M., Murillo, M.L., and Sanchez, O.C. (2009). Beneficial role of dietary folic acid on cholesterol and bile acid metabolism in ethanol-fed rats. *J Stud Alcohol Drugs* 70**,** 615-622.
10. Desantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G.L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72**,** 5069-5072.
11. Dinno, A. (2017). dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums. R package version 1.3.4.
12. Dolédec, S., and Chessel, D. (1994). Co-inertia analysis: an alternative method for studying species–environment relationships. *Freshwater Biology* 31**,** 277-294.
13. Dolédec, S., Chessel, D., Ter Braak, C.J.F., and Champely, S. (1996). Matching species traits to environmental variables: a new three-table ordination method. *Environmental and Ecological Statistics* 3**,** 143-166.
14. Dray, S., Chessel, D., and Thioulouse, J. (2003). CO-INERTIA ANALYSIS AND THE LINKING OF ECOLOGICAL DATA TABLES. *Ecology* 84**,** 3078-3089.
15. Dray, S., and Dufour, A.B. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software* 22**,** 1-20.
16. Dray, S., and Legendre, P. (2008). Testing the species traits-environment relationships: the fourth-corner problem revisited. *Ecology* 89**,** 3400-3412.

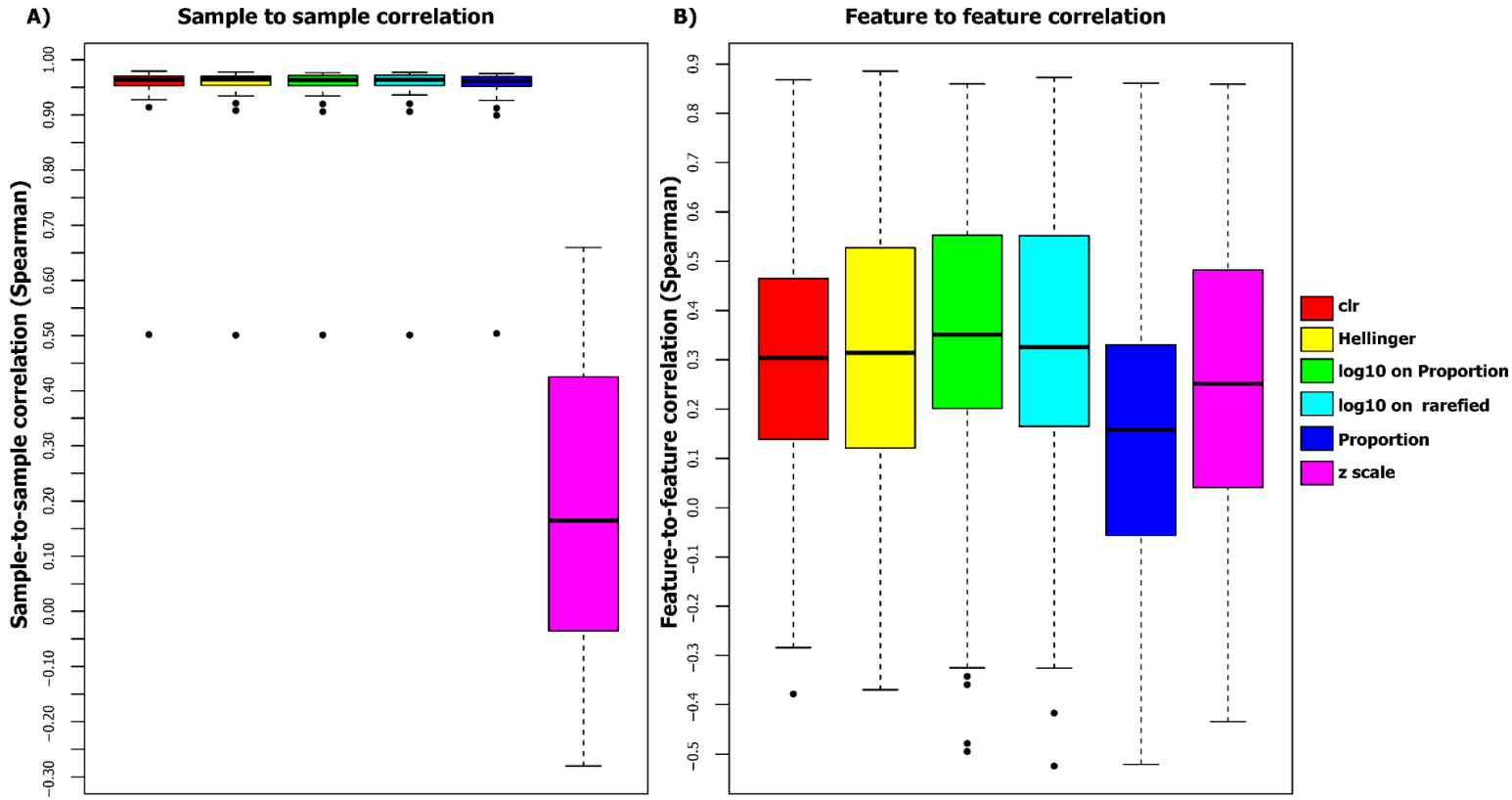
17. Eng, A., and Borenstein, E. (2018). Taxa-function robustness in microbial communities. *Microbiome* 6**,** 45.
18. Franzosa, E.A., Mciver, L.J., Rahnavard, G., Thompson, L.R., Schirmer, M., Weingart, G., Lipson, K.S., Knight, R., Caporaso, J.G., Segata, N., and Huttenhower, C. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. *Nat Methods* 15**,** 962-968.
19. Ghosh, T.S., Das, M., Jeffery, I.B., and O'toole, P.W. (2020). Adjusting for age improves identification of gut microbiome alterations in multiple diseases. *Elife* 9.
20. Hadrich, D. (2018). Microbiome Research Is Becoming the Key to Better Understanding Health and Nutrition. *Front Genet* 9**,** 212.
21. Human Microbiome Project, C. (2012a). A framework for human microbiome research. *Nature* 486**,** 215-221.
22. Human Microbiome Project, C. (2012b). Structure, function and diversity of the healthy human microbiome. *Nature* 486**,** 207-214.
23. Iwai, S., Weinmaier, T., Schmidt, B.L., Albertson, D.G., Poloso, N.J., Dabbagh, K., and Desantis, T.Z. (2016). Piphillin: Improved Prediction of Metagenomic Content by Direct Inference from Human Microbiomes. *PLoS One* 11**,** e0166104.
24. Jovel, J., Patterson, J., Wang, W., Hotte, N., O'keefe, S., Mitchel, T., Perry, T., Kao, D., Mason, A.L., Madsen, K.L., and Wong, G.K. (2016). Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics. *Front Microbiol* 7**,** 459.
25. Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28**,** 27-30.
26. Langille, M.G., Zaneveld, J., Caporaso, J.G., Mcdonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Vega Thurber, R.L., Knight, R., Beiko, R.G., and Huttenhower, C. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* 31**,** 814-821.
27. Legendre, P., and Gallagher, E.D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia* 129**,** 271-280.
28. Neuwirth, E. (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2.
29. Pasolli, E., Schiffer, L., Manghi, P., Renson, A., Obenchain, V., Truong, D.T., Beghini, F., Malik, F., Ramos, M., Dowd, J.B., Huttenhower, C., Morgan, M., Segata, N., and Waldron, L. (2017). Accessible, curated metagenomic data through ExperimentHub. *Nat Methods* 14**,** 1023-1024.
30. Schiffer, L., Azhar, R., Shepherd, L., Ramos, M., Geistlinger, L., Huttenhower, C., Dowd, J.B., Segata, N., and Waldron, L. (2019). HMP16SData: Efficient Access to the Human Microbiome Project Through Bioconductor. *Am J Epidemiol* 188**,** 1023-1026.
31. Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., and Weber, C.F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75**,** 7537-7541.
32. Sze, M.A., Topcuoglu, B.D., Lesniak, N.A., Ruffin, M.T.T., and Schloss, P.D. (2019). Fecal Short-Chain Fatty Acids Are Not Predictive of Colonic Tumor

Status and Cannot Be Predicted Based on Bacterial Community Structure. *MBio* 10.

33. Team, N.I.H.H.M.P.A. (2019). A review of 10 years of human microbiome research activities at the US National Institutes of Health, Fiscal Years 2007-2016. *Microbiome* 7**,** 31.
34. Team, R.C. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing. *Vienna, Austria*.
35. Tessler, M., Neumann, J.S., Afshinnekoo, E., Pineda, M., Hersch, R., Velho, L.F.M., Segovia, B.T., Lansac-Toha, F.A., Lemke, M., Desalle, R., Mason, C.E., and Brugler, M.R. (2017). Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. *Sci Rep* 7**,** 6589.
36. Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York*.
37. Zeller, G., Tap, J., Voigt, A.Y., Sunagawa, S., Kultima, J.R., Costea, P.I., Amiot, A., Bohm, J., Brunetti, F., Habermann, N., Hercog, R., Koch, M., Luciani, A., Mende, D.R., Schneider, M.A., Schrotz-King, P., Tournigand, C., Tran Van Nhieu, J., Yamada, T., Zimmermann, J., Benes, V., Kloor, M., Ulrich, C.M., Von Knebel Doeberitz, M., Sobhani, I., and Bork, P. (2014). Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol* 10**,** 766.

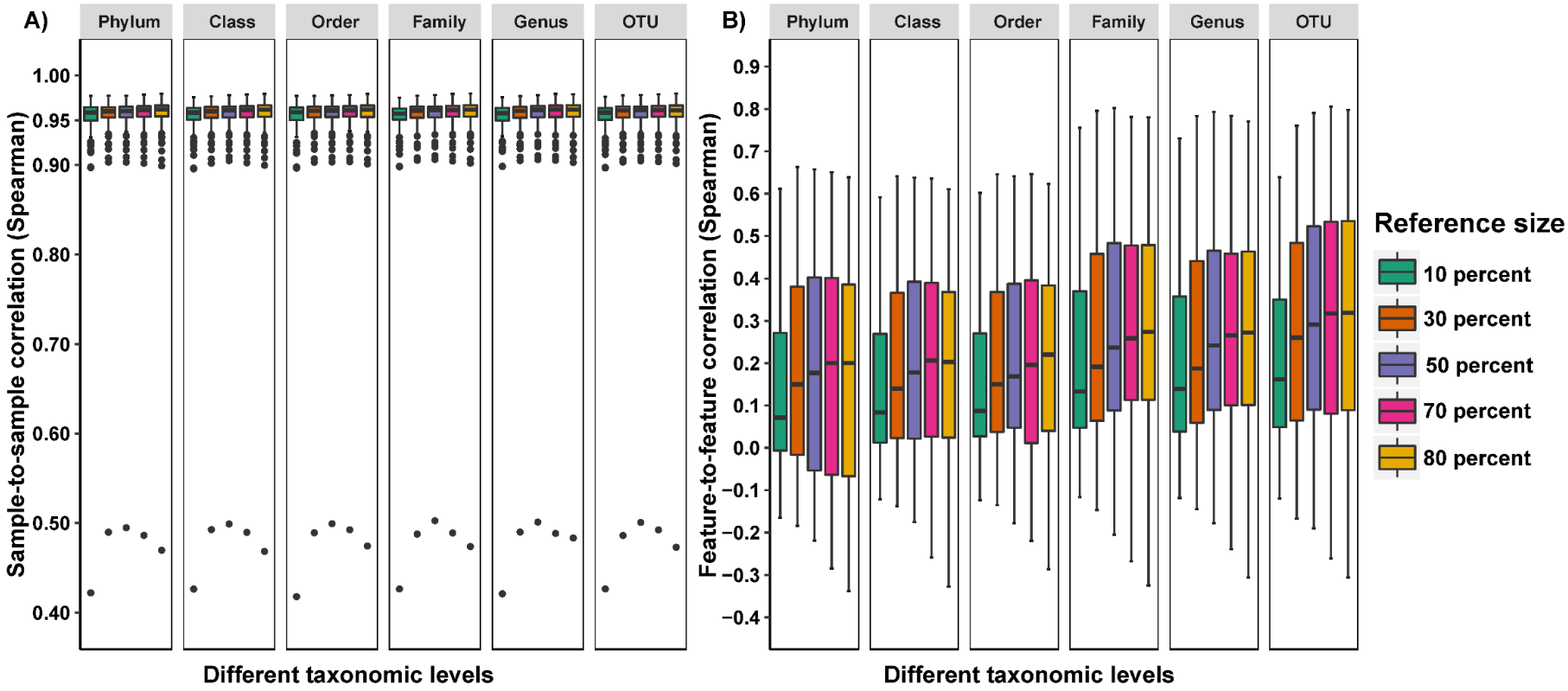
4.9 Supplementary Information

4.9.1 Supplementary Figures



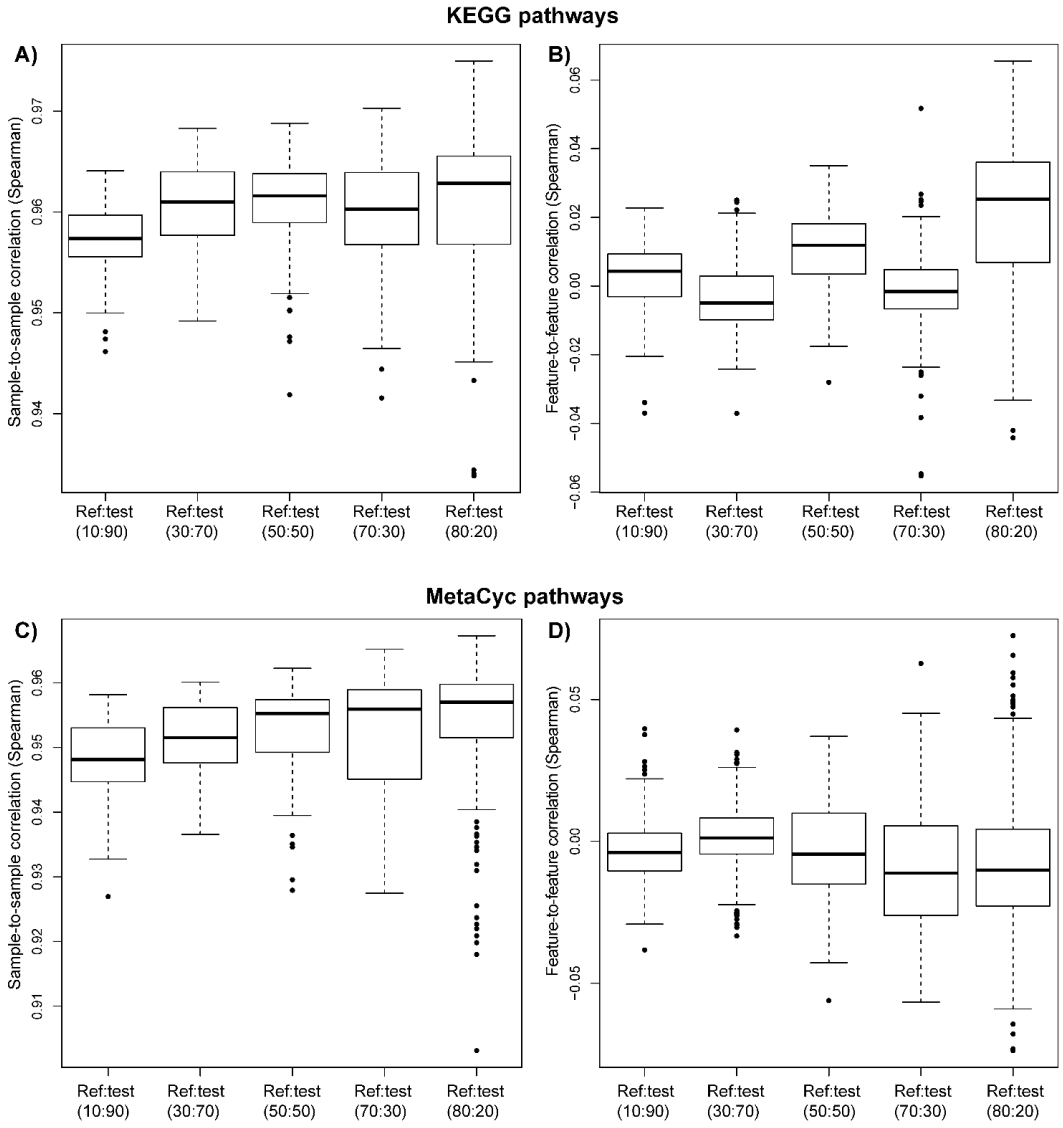
Supplementary figure 1 Sample-to-sample and feature-to-feature correlations between inferred and mWGS functional profiles obtained for MetaCyc pathway schemes using different transformations

The effect of transformation/normalisation methodologies on preliminary analysis of IPCO predictions using the MetaCyc functional profile datasets, in terms of **A)** sample to sample correlations and **B)** the feature to feature correlations for IPCO inferred sample and functional profiles to mWGS sample and functional profiles



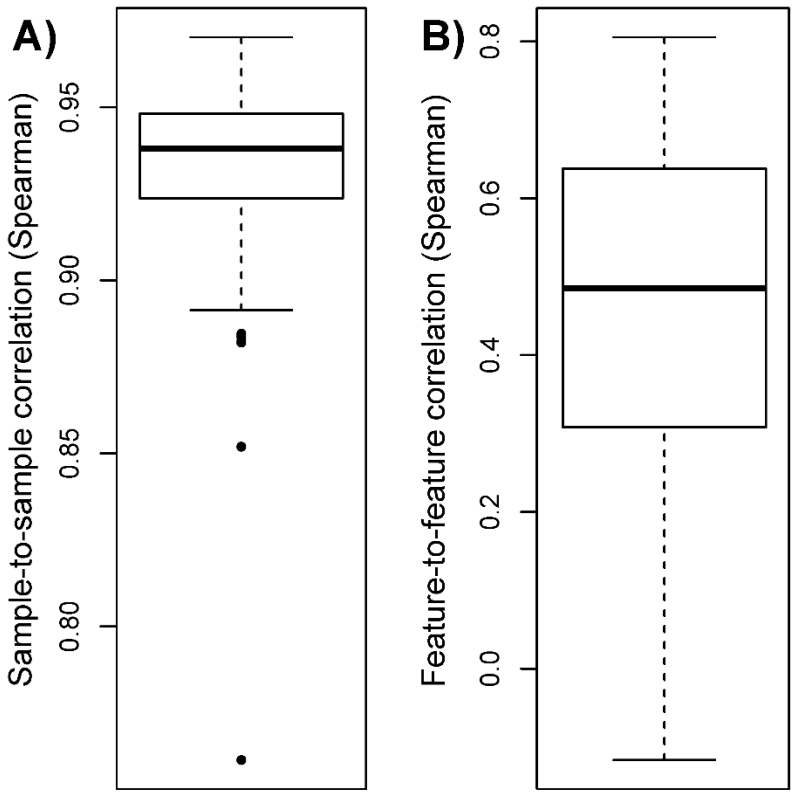
Supplementary figure 2 Comparison of sample-to-sample and feature-to-feature correlations obtained between the inferred and the mWGS MetaCyc pathway abundances at different taxonomic levels and reference dataset size

IPCO's prediction of MetaCyc pathways abundance using different reference dataset size and at different taxonomic levels shows **A)** high sample to sample correlation at all different reference size and taxonomic levels and **B)** shows the correlation values of features improve with larger reference and lowest taxonomic level



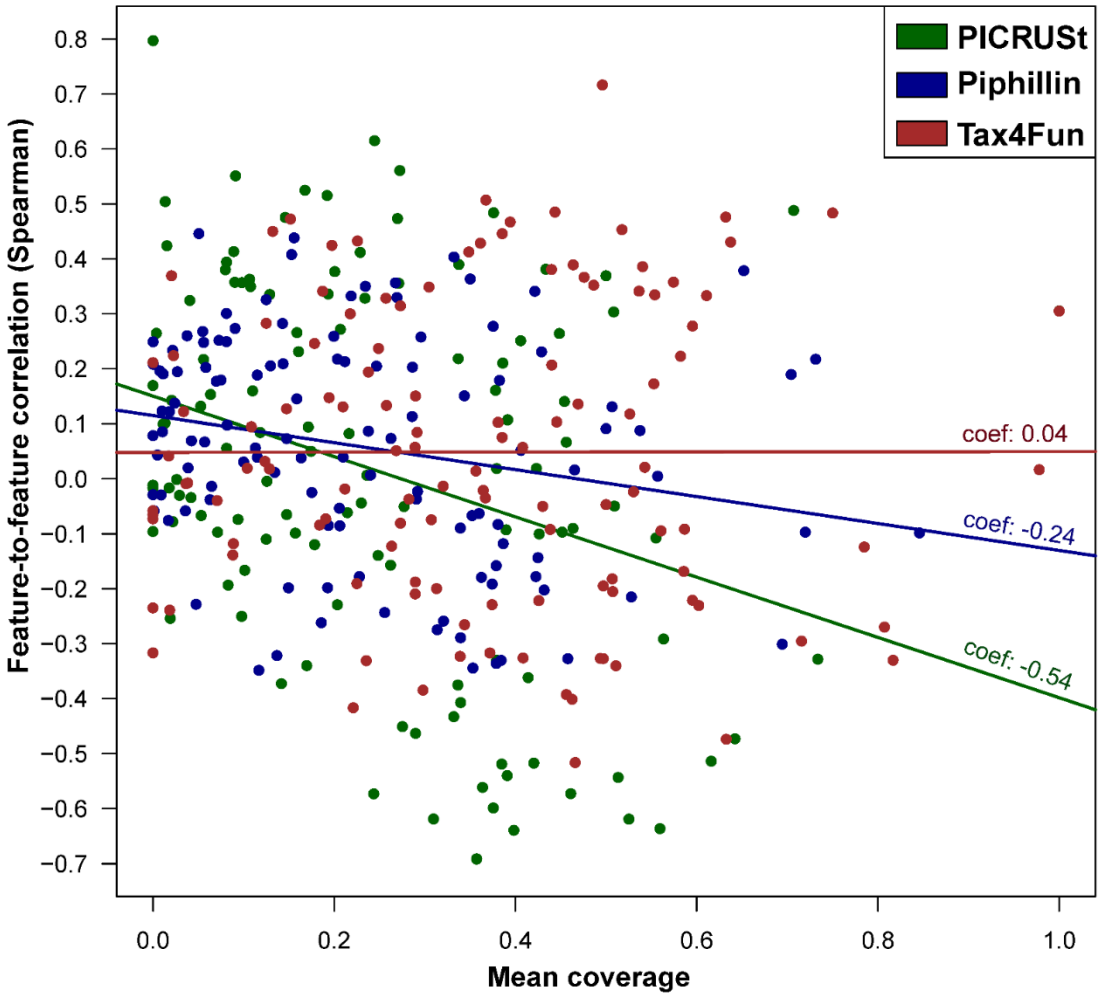
Supplementary figure 3 Comparison of shuffling mWGS samples from query data against inferred profile (KEGG and MetaCyc)

Boxplots showing the correlation values for inferred samples and features against mWGS pathway profiles. Figure 3A-B shows the sample and feature correlation observed with the KEGG pathways profiles when the query sample labels are shuffled. In figure 3C-D, the sample and feature correlation observed between predicted and mWGS MetaCyc profiles



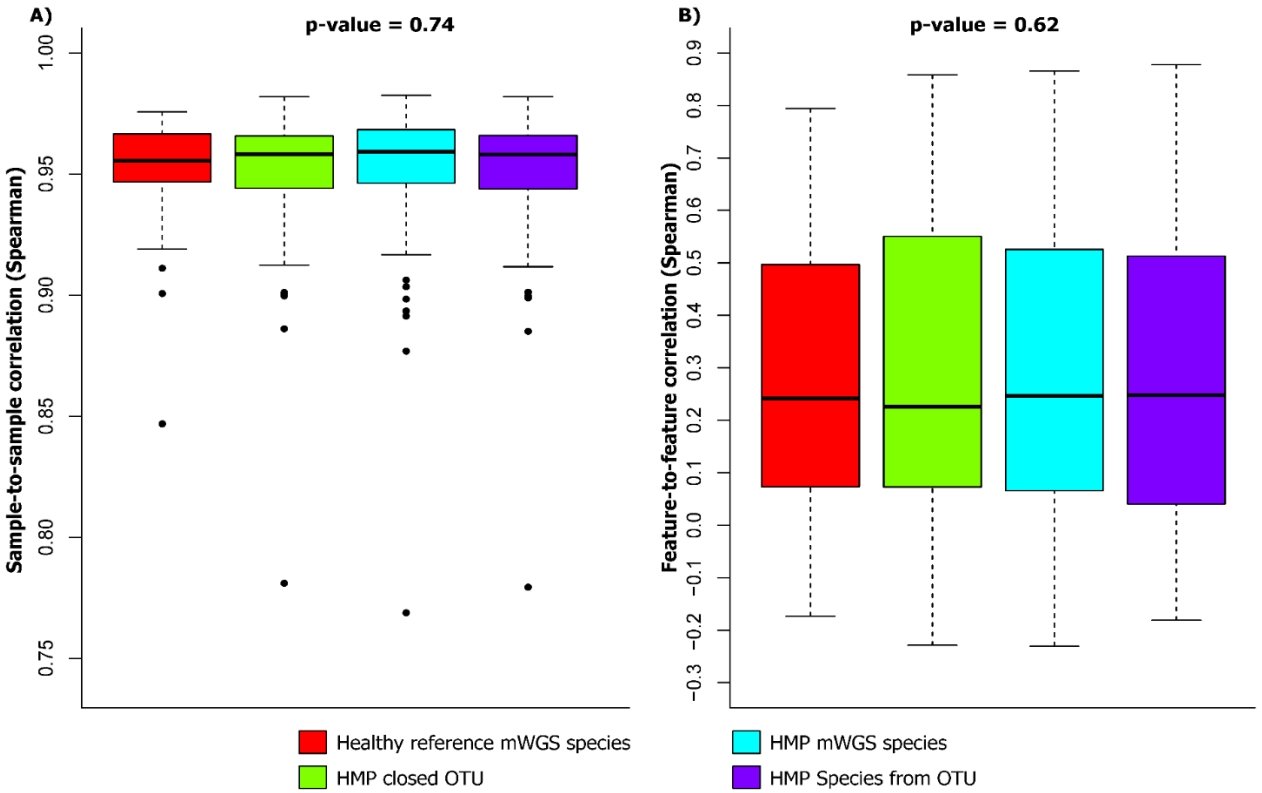
Supplementary figure 4 Evaluating the correlation between IPCO inferred-KO KEGG pathways against the mWGS KEGG pathways

Spearman correlation of KEGG pathways calculated using HUMAnN2 from the IPCO inferred KO profiles compared to mWGS KEGG pathways. A) shows the sample-to-sample correlation observed and B) shows feature-to-feature correlation observed.



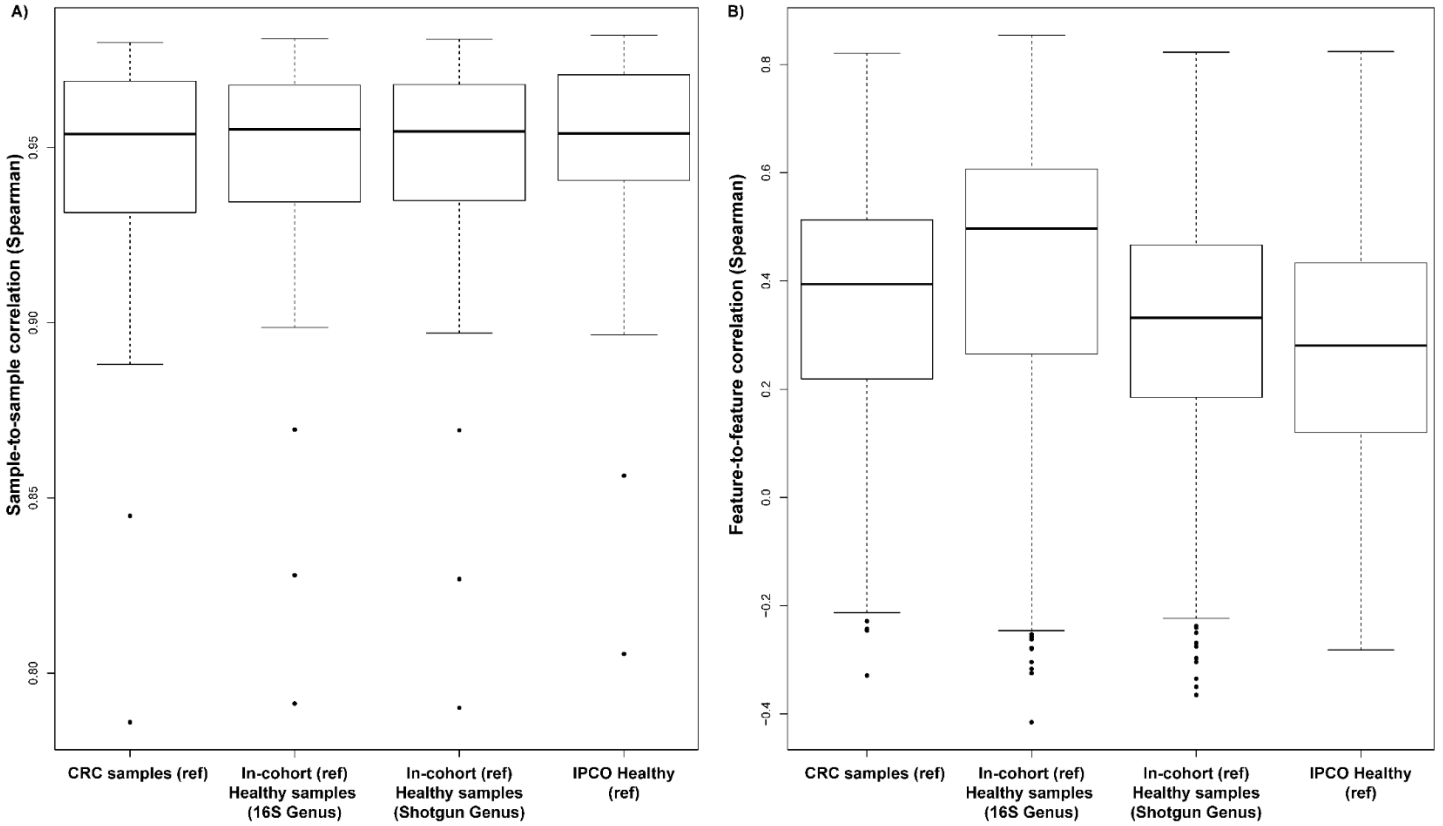
Supplementary figure 5 Feature correlation vs Mean coverage obtained from published tools

A scatter plot of correlation values obtained from feature-to-feature correlation (KEGG pathways) vs mean coverage obtained from the published tools. The coefficient of correlations are reported for the three tools along with the predicted line.



Supplementary figure 6 Comparison of using 16S and mWGS taxonomic dataset as table L when inferring for external dataset

Supplementary figure 6 shows sample and feature correlation of inferred MetaCyc pathways of elderly 16S dataset obtained from using HMP 16S species and closed OTU, HMP mWGS species and healthy reference mWGS species as reference taxonomic dataset (table L). No significant change in observed sample (**A**) and feature (**B**) correlation for the external dataset when using the closed OTU level or species level dataset derived from either 16S representative sequences or mWGS taxonomy (**table L**). The external dataset (**table Q**) is collapsed to closed OTU level or species level derived from 16S representative sequences



Supplementary figure 7 Accuracy of using different references to predict MetaCyc profiles for CRC samples

Supplementary figure 7 highlights the sample (A) and feature (B) correlation obtained when using different samples types as reference taxonomy and functional datasets to predict MetaCyc profiles from 16S genus level of only CRC samples.

CRC samples (ref): Using the CRC samples as references, In-cohort (ref) Healthy samples (16S Genus): Using the healthy samples from the same cohort with 16S Genus profiles as reference taxonomy, In-cohort (ref) Healthy samples (Shotgun Genus): Reference is healthy samples from the same cohort with refernce mWGS Genus profiles, IPCO Healthy (ref): The healthy samples provided with IPCO as references after excluding Zeller *et al.* samples

4.9.2 Supplementary Tables

Supplementary table 1 Sample-to-sample and feature-to-feature covariance observed between mWGS and IPCO inferred functional profiles using different normalisation/transformation

Normalisation/transformation	Sample-to-sample covariance		Feature-to-feature covariance	
	RV	P-value	RV	P-value
Hellinger	0.89	0.001	0.54	0.007
Log10 on rarefied counts	0.84	0.001	0.33	0.003
Log10 on proportion	0.84	0.001	0.31	0.002
Proportion	0.97	0.001	0.36	0.086
z-scaling	0.24	0.001	0.20	0.079
clr transformation	0.28	0.011	0.55	0.001

Comparison of the effect of different normalisation/transformation on the sample to sample and feature to feature co-inertia between IPCO inferred and mWGS MetaCyc pathways abundance datasets. Co-variance is determined by RV coefficient. All observed co-variance were significant at nominal p-value ≤ 0.05 except for proportional and z-scaled features co-variance. Number of samples in reference dataset size is 70% more than the queried dataset.

Supplementary table 2 Pairwise comparison of correlations observed using reference datasets of different size and taxonomic level

A) Significance (P-values) of the correlations between mWGS and IPCO-inferred KEGG pathway abundance profiles

Samples						
Reference size	**Phylum**	**Class**	**Order**	**Family**	**Genus**	**OTU**
10 - 30	2.6E-01	1.8E-01	3.1E-01	3.9E-01	3.0E-01	4.1E-01
10 - 50	2.1E-01	1.4E-01	2.1E-01	2.6E-01	2.6E-01	2.6E-01
10 - 70	2.2E-01	2.9E-01	5.0E-01	3.5E-01	6.7E-01	6.1E-01
10 - 80	4.1E-01	1.8E-01	2.5E-01	6.8E-01	3.6E-01	3.6E-01
30 - 50	5.6E-01	5.5E-01	5.0E-01	4.9E-01	5.7E-01	4.8E-01
30 - 70	5.5E-01	6.6E-01	6.2E-01	5.3E-01	7.4E-01	5.7E-01
30 - 80	6.3E-01	6.1E-01	5.1E-01	6.3E-01	6.5E-01	5.2E-01
50 - 70	4.8E-01	5.5E-01	5.7E-01	5.3E-01	5.9E-01	5.6E-01
50 - 80	5.2E-01	4.8E-01	5.1E-01	5.9E-01	4.9E-01	4.8E-01
70 - 80	4.9E-01	5.1E-01	5.0E-01	4.9E-01	5.4E-01	5.2E-01
Features						
Reference size	**Phylum**	**Class**	**Order**	**Family**	**Genus**	**OTU**
10 - 30	**1.2E-02**	**8.3E-04**	**7.2E-04**	**1.2E-02**	**3.4E-02**	**1.3E-05**
10 - 50	**3.7E-03**	**3.4E-05**	**2.3E-06**	**2.5E-03**	**2.6E-03**	**1.7E-08**
10 - 70	**2.9E-03**	**1.1E-04**	**6.2E-07**	**1.4E-03**	**3.1E-03**	**3.8E-09**
10 - 80	**4.2E-03**	**9.7E-05**	**3.0E-07**	**1.7E-02**	**1.3E-02**	**1.2E-09**
30 - 50	4.7E-01	2.7E-01	1.2E-01	3.1E-01	2.1E-01	1.3E-01
30 - 70	4.3E-01	4.1E-01	7.2E-02	2.8E-01	2.6E-01	8.5E-02
30 - 80	4.6E-01	3.8E-01	**4.8E-02**	4.6E-01	3.6E-01	5.4E-02
50 - 70	4.8E-01	3.7E-01	4.1E-01	4.8E-01	4.0E-01	4.2E-01
50 - 80	4.8E-01	3.9E-01	3.5E-01	3.1E-01	3.0E-01	3.4E-01
70 - 80	5.2E-01	4.8E-01	4.0E-01	2.8E-01	3.6E-01	3.8E-01

Significance of correlation values of IPCO inferred samples and features obtained through pairwise comparison between different KEGG reference datasets size. Significance is determined by P-adjusted ≤ 0.05 and highlighted in red and bold

B) Significance (P-values) of the correlations between mWGS and IPCO-inferred MetaCyc pathway abundance profiles

Samples						
Reference size	**Phylum**	**Class**	**Order**	**Family**	**Genus**	**OTU**
10 - 30	3.8E-01	2.9E-01	4.8E-01	8.1E-02	2.2E-01	1.6E-01
10 - 50	4.4E-01	1.4E-01	4.2E-01	**2.5E-02**	7.4E-02	1.5E-01
10 - 70	4.4E-01	1.2E-01	4.0E-01	**1.9E-02**	**4.9E-02**	8.6E-02
10 - 80	4.1E-01	1.4E-01	4.0E-01	**1.9E-02**	5.9E-02	1.2E-01
30 - 50	3.8E-01	4.2E-01	4.4E-01	4.0E-01	3.7E-01	4.9E-01
30 - 70	3.8E-01	3.6E-01	4.2E-01	3.4E-01	2.7E-01	4.7E-01
30 - 80	4.4E-01	3.2E-01	3.8E-01	2.9E-01	2.5E-01	4.6E-01
50 - 70	4.2E-01	4.5E-01	4.0E-01	4.1E-01	4.1E-01	4.2E-01
50 - 80	4.2E-01	4.0E-01	4.5E-01	4.0E-01	3.8E-01	4.0E-01
70 - 80	3.9E-01	4.2E-01	4.6E-01	4.5E-01	4.3E-01	4.4E-01
Features						
Reference size	**Phylum**	**Class**	**Order**	**Family**	**Genus**	**OTU**
10 - 30	**2.2E-03**	**2.9E-05**	**2.3E-05**	**4.7E-05**	**3.3E-05**	**2.6E-08**
10 - 50	**1.9E-02**	**1.1E-06**	**5.9E-08**	**7.6E-09**	**3.9E-11**	**6.0E-16**
10 - 70	**1.7E-02**	**2.5E-07**	**9.6E-06**	**2.8E-11**	**4.0E-13**	**1.7E-16**
10 - 80	7.1E-02	**4.7E-06**	**5.8E-08**	**2.8E-11**	**3.3E-13**	**3.2E-17**
30 - 50	2.8E-01	3.4E-01	1.6E-01	5.9E-02	**9.4E-03**	**7.7E-03**
30 - 70	2.9E-01	2.2E-01	4.5E-01	**6.4E-03**	**1.2E-03**	**5.0E-03**
30 - 80	1.1E-01	4.1E-01	1.6E-01	**5.7E-03**	**1.0E-03**	**2.5E-03**
50 - 70	4.6E-01	3.8E-01	1.8E-01	2.0E-01	2.7E-01	4.2E-01
50 - 80	2.8E-01	3.6E-01	4.5E-01	1.9E-01	2.6E-01	4.0E-01
70 - 80	2.8E-01	3.2E-01	1.7E-01	4.6E-01	4.5E-01	4.3E-01

Significance of correlation values of IPCO inferred samples and features obtained through pairwise comparison between different MetaCyc reference datasets size. Significance is determined by P-adjusted ≤ 0.05 and highlighted in red and bold

Supplementary table 3 Comparison of sample-to-sample and feature-to-feature correlations between the inferred and the mWGS pathway profiles (KEGG pathways) obtained using different methodologies for different sites

Samples					
	Water (Brazilian river)	**nasal**	**oral**	**skin**	**stool**
IPCO - PICRUSt	**5.5E-10**	**4.4E-11**	**3.0E-13**	**7.8E-03**	**9.2E-16**
IPCO - Piphillin	**1.7E-10**	**8.2E-10**	**3.0E-13**	**7.8E-03**	**9.2E-16**
IPCO - Tax4Fun	**5.5E-10**	**2.3E-11**	**3.0E-13**	**7.8E-03**	**9.2E-16**
PICRUSt - Piphillin	**2.5E-07**	**3.4E-09**	**4.1E-03**	**7.8E-03**	**1.3E-03**
PICRUSt - Tax4Fun	7.3E-01	**2.3E-11**	**3.0E-13**	**7.8E-03**	**9.2E-16**
Piphillin - Tax4Fun	**3.8E-07**	**2.3E-11**	**3.0E-13**	**7.8E-03**	**3.5E-15**
Features					
	Water (Brazilian river)	**nasal**	**oral**	**skin**	**stool**
IPCO - PICRUSt	**2.0E-08**	**1.4E-05**	**3.7E-06**	**1.4E-05**	**2.6E-16**
IPCO - Piphillin	**3.5E-10**	**2.4E-07**	**8.0E-09**	**7.9E-09**	**2.6E-17**
IPCO - Tax4Fun	**4.3E-08**	**2.9E-07**	**5.7E-08**	**1.5E-09**	**1.8E-17**
PICRUSt - Piphillin	7.1E-01	8.2E-01	4.1E-01	**3.3E-02**	1.2E-01
PICRUSt - Tax4Fun	6.4E-01	4.2E-01	6.3E-01	**1.1E-05**	**4.6E-02**
Piphillin - Tax4Fun	7.1E-01	8.6E-01	6.3E-01	**7.0E-03**	3.6E-01

Pairwise comparison of sample-to-sample and feature-to-feature correlation values (KEGG pathways) from different methodology. Significance is determined by P-adjusted ≤ 0.05 and highlighted in red and bold.

Supplementary table 4 Comparison of sample-to-sample and feature-to-feature correlations between the inferred and the mWGS KEGG Orthologs (KO) profiles obtained using different tools for different sites

Samples					
	Water (Brazilian river)	**nasal**	**oral**	**skin**	**stool**
IPCO - PICRUSt	**8.1E-08**	**2.3E-09**	**7.8E-03**	**8.6E-16**	**3.2E-13**
IPCO - Piphillin	**4.4E-10**	**8.7E-08**	**7.8E-03**	**8.6E-16**	**3.2E-13**
IPCO - Tax4Fun	**2.6E-09**	**7.1E-11**	**7.8E-03**	**8.6E-16**	**3.2E-13**
PICRUSt - Piphillin	**7.3E-06**	**6.8E-10**	**7.8E-03**	**2.8E-15**	**3.2E-13**
PICRUSt - Tax4Fun	**8.7E-11**	**3.4E-11**	**7.8E-03**	**8.6E-16**	**3.2E-13**
Piphillin - Tax4Fun	7.5E-02	**3.4E-11**	**7.8E-03**	9.6E-01	**3.2E-13**
Features					
	Water (Brazilian river)	**nasal**	**oral**	**skin**	**stool**
IPCO - PICRUSt	**8.0E-24**	3.2E-01	**5.5E-107**	**2.0E-04**	**1.4E-52**
IPCO - Piphillin	**1.5E-96**	**9.7E-03**	**2.1E-97**	6.0E-02	**3.7E-57**
IPCO - Tax4Fun	**1.9E-38**	**1.4E-04**	**2.2E-90**	2.5E-01	**9.8E-260**
PICRUSt - Piphillin	**3.5E-48**	**2.8E-04**	**1.0E-05**	**3.2E-27**	**7.8E-03**
PICRUSt - Tax4Fun	**1.5E-06**	**3.3E-06**	**1.8E-03**	**3.8E-04**	**1.4E-96**
Piphillin - Tax4Fun	**6.7E-32**	**4.0E-02**	**3.1E-02**	**2.3E-05**	**1.8E-49**

Pairwise comparison of sample-to-sample and feature-to-feature correlation values using KO profiles from different methodology. Significance is determined by P-adjusted ≤ 0.05 and highlighted in red and bold.

Supplementary table 5 Co-variance observed between taxonomic and functional datasets from different sites

KEGG				
	OTU & Pathway abundance		mWGS species & Pathway abundance	
site	RV	pvalue	RV	pvalue
nasal	0.08	0.892	0.07	0.404
oral	0.14	0.845	0.12	0.366
skin	0.47	0.795	0.43	0.390
stool	0.35	**0.001**	0.35	**0.001**
Water (Brazilian river)	0.17	0.188	0.23	**0.009**
MetaCyc				
site	RV	pvalue	RV	pvalue
nasal	0.11	0.945	0.07	0.745
oral	0.14	0.911	0.13	0.233
skin	0.43	0.929	0.36	0.630
stool	0.41	0.001	0.47	0.001
Water (Brazilian river)	0.19	0.146	0.22	**0.012**

Co-inertia of samples and pathways between the inferred KEGG/MetaCyc profiles with its paired mWGS pathways abundance. Degree of co-variance is determined by RV coefficient and significance is determined by nominal p-value ≤ 0.05 and highlighted in red and bold

Supplementary table 6 Comparison of inferred vs mWGS profiles of CRC samples
A) Pairwise comparison between inferred and mWGS KEGG pathway profiles obtained using various references implemented in IPCO and by using published tools

Pairwise comparisons	P-values (Sample-to-sample)	P-values (Feature-to-feature)
CRC samples (ref) vs In-cohort (ref) Healthy samples (16S Genus)	**3.2E-02**	**2.7E-08**
CRC samples (ref) vs In-cohort (ref) Healthy samples (Shotgun Genus)	**2.3E-02**	**4.6E-02**
CRC samples (ref) vs IPCO Healthy (ref)	7.2E-01	**1.1E-08**
CRC samples (ref) vs PICRUSt	**7.2E-12**	**6.0E-11**
CRC samples (ref) vs Piphillin	**1.1E-11**	**3.6E-10**
CRC samples (ref) vs Tax4Fun	**3.2E-12**	**2.2E-17**
In-cohort (ref) Healthy samples (16S Genus) vs In-cohort (ref) Healthy samples (Shotgun Genus)	**1.3E-04**	**2.3E-14**
In-cohort (ref) Healthy samples (16S Genus) vs IPCO Healthy (ref)	6.0E-01	3.6E-01

In-cohort (ref) Healthy samples (16S Genus) vs PICRUSt	**1.9E-11**	**4.4E-15**
In-cohort (ref) Healthy samples (16S Genus) vs Piphillin	**1.1E-11**	**3.3E-14**
In-cohort (ref) Healthy samples (16S Genus) vs Tax4Fun	**3.2E-12**	**7.7E-18**
In-cohort (ref) Healthy samples (Shotgun Genus) vs IPCO Healthy (ref)	6.0E-01	**1.4E-12**
In-cohort (ref) Healthy samples (Shotgun Genus) vs PICRUSt	**1.9E-11**	**5.5E-10**
In-cohort (ref) Healthy samples (Shotgun Genus) vs Piphillin	**1.1E-11**	**5.1E-10**
In-cohort (ref) Healthy samples (Shotgun Genus) vs Tax4Fun	**3.2E-12**	**1.6E-16**
IPCO Healthy (ref) vs PICRUSt	**5.5E-12**	**1.6E-15**
IPCO Healthy (ref) vs Piphillin	**1.1E-11**	**1.1E-14**
IPCO Healthy (ref) vs Tax4Fun	**3.2E-12**	**2.8E-18**
PICRUSt vs Piphillin	4.1E-01	4.9E-01
PICRUSt vs Tax4Fun	**3.2E-12**	**2.1E-05**
Piphillin vs Tax4Fun	**3.2E-12**	**6.1E-04**

Significance of all pairwise comparisons when inferring KEGG pathways functionality of colorectal cancer (CRC) samples using various references implemented in IPCO along with predictions from PICRUSt, Piphillin and Tax4Fun. Significant p-values (P-adjusted ≤ 0.05) is highlighted in red and bold.

B) Pairwise comparison between inferred and mWGS KEGG pathway profiles obtained using various references implemented in IPCO

Pairwise comparisons	P-values (Sample-to-sample)	P-values (Feature-to-feature)
In-cohort (ref) Healthy samples (16S Genus) vs CRC samples (ref)	5.0E-01	**3.8E-45**
In-cohort (ref) Healthy samples (Shotgun Genus) vs CRC samples (ref)	5.0E-01	**1.9E-14**
IPCO Healthy (ref) vs CRC samples (ref)	5.0E-01	**3.8E-14**
In-cohort (ref) Healthy samples (Shotgun Genus) vs In-cohort (ref) Healthy samples (16S Genus)	5.0E-01	**3.8E-75**
IPCO Healthy (ref) vs In-cohort (ref) Healthy samples (16S Genus)	5.0E-01	**2.2E-45**
IPCO Healthy (ref) vs In-cohort (ref) Healthy samples (Shotgun Genus)	5.0E-01	**3.2E-04**

Significance of all pairwise comparisons when inferring MetaCyc pathways functionality of colorectal cancer (CRC) samples using various references implemented in IPCO. Significant p-values (P-adjusted ≤ 0.05) is highlighted in red and bold.

Supplementary table 7 Correlation between the paired bile acid profiles and the different KEGG pathways obtained using mWGS and those inferred using different methods

Spearman correlation values						
Bile acids	**KEGG pathways**	**Shotgun**	**IPCO**	**PICRUSt**	**Tax4Fun**	**Piphillin**
cholic acid	ko00121: Secondary bile acid biosynthesis	-0.34	-0.21	0.37	-0.08	0.00
cholic acid	ko00790: Folate biosynthesis	0.48	0.36	0.02	0.34	0.36
Chenodeoxycholic acid	ko00121: Secondary bile acid biosynthesis	-0.27	-0.22	0.31	-0.07	-0.05
Chenodeoxycholic acid	ko00790: Folate biosynthesis	0.43	0.31	0.06	0.36	0.35
Lithocholic acid	ko00430: Taurine and hypotaurine metabolism	0.35	0.22	0.11	0.22	0.3
Lithocholic acid	ko03070: Bacterial secretion system	0.37	0.2	-0.15	-0.23	-0.01
Dehydrocholic acid	ko05100: Bacterial invasion of epithelial cells	-0.27	-0.21	-0.30	-0.20	-0.42
12-Ketolithocholic acid	ko00121: Secondary bile acid biosynthesis	0.24	-0.12	0.08	-0.05	0.06
12-Ketolithocholic acid	ko00430: Taurine and hypotaurine metabolism	0.24	0.03	-0.08	0.05	0.07
dehydrolithocholic acid	ko00121: Secondary bile acid biosynthesis	0.25	0.13	-0.11	-0.01	0.19
dehydrolithocholic acid	ko00430: Taurine and hypotaurine metabolism	0.28	-0.08	0.22	-0.02	0.14
7-ketolithocholic acid	ko00790: Folate biosynthesis	0.003	0.006	-0.05	0.001	0.021
Hyodeoxycholic acid	ko00430: Taurine and hypotaurine metabolism	-0.27	-0.19	0.04	-0.10	-0.08
Ursodeoxycholic acid	ko00790: Folate biosynthesis	0.49	0.42	0.01	0.38	0.36
Dioxolithocholic acid	ko00790: Folate biosynthesis	0.32	0.18	0.15	0.19	0.15
Isolithocholic acid	ko00430: Taurine and hypotaurine metabolism	0.27	0.04	0.17	0.17	0.25
P-values of Spearman correlation						
cholic acid	ko00121: Secondary bile acid biosynthesis	**0.014**	0.064	**0.001**	0.497	0.971
cholic acid	ko00790: Folate biosynthesis	**0.001**	**0.001**	0.865	**0.002**	**0.001**
Chenodeoxycholic acid	ko00121: Secondary bile acid biosynthesis	**0.052**	**0.049**	**0.005**	0.521	0.638
Chenodeoxycholic acid	ko00790: Folate biosynthesis	**0.002**	**0.006**	0.605	**0.001**	**0.001**
Lithocholic acid	ko00430: Taurine and hypotaurine metabolism	**0.013**	**0.049**	0.340	0.051	**0.007**
Lithocholic acid	ko03070: Bacterial secretion system	**0.008**	0.08	0.180	**0.04**	0.927
Dehydrocholic acid	ko05100: Bacterial invasion of epithelial cells	**0.053**	0.059	**0.006**	0.084	**0.000**
12-Ketolithocholic acid	ko00121: Secondary bile acid biosynthesis	**0.099**	0.293	0.492	0.68	0.573
12-Ketolithocholic acid	ko00430: Taurine and hypotaurine metabolism	**0.085**	0.809	0.472	0.647	0.541
dehydrolithocholic acid	ko00121: Secondary bile acid biosynthesis	**0.072**	0.251	0.326	0.897	0.089
dehydrolithocholic acid	ko00430: Taurine and hypotaurine metabolism	**0.048**	0.462	0.056	0.842	0.207
7-ketolithocholic acid	ko00790: Folate biosynthesis	**0.003**	**0.006**	0.673	**0.001**	**0.021**
Hyodeoxycholic acid	ko00430: Taurine and hypotaurine metabolism	**0.051**	0.089	0.739	0.360	0.479
Ursodeoxycholic acid	ko00790: Folate biosynthesis	**0.001**	**0.0001**	0.943	**0.001**	**0.001**
Dioxolithocholic acid	ko00790: Folate biosynthesis	**0.02**	0.123	**0.191**	0.098	0.178
Isolithocholic acid	ko00430: Taurine and hypotaurine metabolism	**0.055**	0.757	0.131	0.138	**0.025**

Correlation between mWGS key KEGG pathways profiles with its paired bile acid metabolites levels and replication of correlations with inferred profiles from different methodology with bile acids levels. Significance for the mWGS profile is determined by p-adjusted ≤ 0.1 and for the different methodology by nominal p-value ≤ 0.1 and highlighted in red and bold. Directionality and strength of correlation is determined by spearman correlation estimate.

Supplementary table 8 Correlation between mWGS and inferred MetaCyc pathway pathway abundance and their paired bile acid profiles

		Shotgun			IPCO	
Bile_acids	**MetaCyc pathways**	**cor value**	**pvalue**	**padj**	**cor value**	**pval**
cholic acid	1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis	0.30	**0.007**	**0.029**	0.30	**0.008**
cholic acid	PWY-6518: glycocholate metabolism (bacteria)	0.32	**0.005**	**0.019**	0.37	**0.001**
Chenodeoxycholic acid	1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis	0.28	**0.013**	**0.051**	0.25	**0.028**
Chenodeoxycholic acid	PWY-6518: glycocholate metabolism (bacteria)	0.28	**0.012**	**0.048**	0.33	**0.003**
Deoxycholic acid	PWY-6518: glycocholate metabolism (bacteria)	0.33	**0.003**	**0.030**	0.26	**0.022**
Lithocholic acid	1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis	0.31	**0.006**	0.201	0.22	**0.048**
Lithocholic acid	PWY-6518: glycocholate metabolism (bacteria)	0.27	**0.018**	0.234	0.27	**0.017**
7-ketolithocholic acid	1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis	0.30	**0.008**	**0.052**	0.24	**0.036**
7-ketolithocholic acid	PWY-6518: glycocholate metabolism (bacteria)	0.26	**0.020**	**0.086**	0.33	**0.003**
Ursodeoxycholic acid	1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis	0.31	**0.006**	**0.021**	0.40	**0.0002**
Ursodeoxycholic acid	PWY-6518: glycocholate metabolism (bacteria)	0.27	**0.017**	**0.053**	0.38	**0.001**
Dioxolithocholic acid	1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis	0.24	**0.032**	0.115	0.19	**0.099**
Isolithocholic acid	1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis	0.23	**0.046**	0.990	0.06	0.600

Correlation between mWGS key MetaCyc pathways profiles with its paired bile acid metabolites levels and replication of correlations between IPCO inferred profiles with bile acids levels. Significance for the mWGS profile is determined by nominal p-value ≤ 0.1 and for IPCO profiles by nominal p-value ≤ 0.1. The p-adj values (≤ 0.1) are also noted for the mWGS results. All significances are highlighted in red and bold. Directionality and strength of correlation is determined by spearman correlation estimate.

Supplementary table 9 Correlation of mWGS and inferred KEGG pathways to its paired butyrate and propionate levels

Spearman correlation values

			shotgun	IPCO	PICRUSt	Tax4Fun	Piphillin
Butyrate	ko00250:	Alanine, aspartate and glutamate metabolism	0.06	0.02	0.06	0.06	-0.22
	ko00300:	Lysine biosynthesis	0.17	0.03	0.06	-0.03	-0.24
	ko00310:	Lysine degradation	-0.09	-0.13	0.03	0.04	-0.10
	ko00471:	D-Glutamine and D-glutamate metabolism	0.08	-0.02	0.02	0.22	-0.12
	ko00650:	Butanoate metabolism	0.12	-0.11	0.08	-0.05	-0.19
	ko04974:	Protein digestion and absorption	-0.07	-0.02	0.03	0.09	0.04
Propionate	ko00250:	Alanine, aspartate and glutamate metabolism	0.07	0.01	0.07	0.10	-0.15
	ko00260:	Glycine, serine and threonine metabolism	0.11	0.01	-0.01	0.08	-0.17
	ko00270:	Cysteine and methionine metabolism	0.12	-0.18	-0.03	-0.14	-0.06
	ko00340:	Histidine metabolism	0.10	0.02	-0.14	0.00	-0.30
	ko00471:	D-Glutamine and D-glutamate metabolism	0.08	-0.09	0.00	0.30	-0.05
	ko00473:	D-Alanine metabolism	0.04	-0.31	0.01	-0.14	0.01
	ko00640:	Propanoate metabolism	-0.11	-0.20	-0.11	-0.17	-0.14
	ko04974:	Protein digestion and absorption	-0.05	0.00	0.10	0.17	0.17
P-values of Spearman correlation							
Butyrate	ko00250:	Alanine, aspartate and glutamate metabolism	0.59	0.83	0.58	0.57	**0.06**
	ko00300:	Lysine biosynthesis	0.13	0.76	0.59	0.79	**0.04**
	ko00310:	Lysine degradation	0.44	0.24	0.76	0.74	0.39
	ko00471:	D-Glutamine and D-glutamate metabolism	0.48	0.83	0.86	**0.05**	0.31
	ko00650:	Butanoate metabolism	0.29	0.32	0.46	0.68	**0.09**
	ko04974:	Protein digestion and absorption	0.56	0.87	0.80	0.43	0.70
Propionate	ko00250:	Alanine, aspartate and glutamate metabolism	0.52	0.90	0.55	0.37	0.19
	ko00260:	Glycine, serine and threonine metabolism	0.35	0.95	0.93	0.50	0.14
	ko00270:	Cysteine and methionine metabolism	0.31	0.12	0.77	0.23	0.57
	ko00340:	Histidine metabolism	0.39	0.89	0.21	0.99	**0.01**
	ko00471:	D-Glutamine and D-glutamate metabolism	0.47	0.43	0.97	**0.01**	0.64
	ko00473:	D-Alanine metabolism	0.72	**0.01**	0.96	0.22	0.93
	ko00640:	Propanoate metabolism	0.35	**0.07**	0.32	0.13	0.23
	ko04974:	Protein digestion and absorption	0.68	1.00	0.39	0.13	0.14

Correlation between mWGS key KEGG pathways profiles with its paired butyrate and propionate metabolites levels and replication of correlations between inferred profiles from different methodology with butyrate and propionate levels. Significance for the mWGS profile and for the different methodology by nominal p-value ≤ 0.1, highlighted in red, and bold. mWGS profiles were not significantly correlated with butyrate and propionate levels. Directionality and strength of correlation is determined by spearman correlation estimate.

Supplementary table 10 Estimates and p-values of correlation from key mWGS and inferred MetaCyc pathways to its paired butyrate and propionate levels

SCFAs	MetaCyc pathways	shotgun			IPCO	
		cor value	pval	padj	cor value	pval
Butyrate	ARGDEG-PWY: superpathway of L-arginine, putrescine, and 4-aminobutanoate degradation	0.041	0.721	0.963	0.036	0.751
	CENTFERM-PWY: pyruvate fermentation to butanoate	0.052	0.650	0.936	-0.118	0.298
	DAPLYSINESYN-PWY: L-lysine biosynthesis I	-0.030	0.791	0.963	-0.149	0.190
	P163-PWY: L-lysine fermentation to acetate and butanoate	-0.079	0.487	0.883	-0.206	**0.069**
	P4-PWY: superpathway of L-lysine, L-threonine and L-methionine biosynthesis I	-0.168	0.139	0.526	-0.146	0.200
	PWY-2941: L-lysine biosynthesis II	-0.018	0.873	0.979	-0.053	0.643
	PWY-2942: L-lysine biosynthesis III	0.248	**0.028**	0.230	0.070	0.540
	PWY-5022: 4-aminobutanoate degradation V	-0.043	0.704	0.963	-0.187	**0.098**
	PWY-5097: L-lysine biosynthesis VI	0.350	**0.002**	0.125	0.149	0.191
	PWY-5100: pyruvate fermentation to acetate and lactate II	0.017	0.882	0.979	-0.079	0.488
	PWY-5505: L-glutamate and L-glutamine biosynthesis	0.097	0.395	0.820	-0.147	0.197
	PWY-5676: acetyl-CoA fermentation to butanoate II	-0.082	0.473	0.867	-0.016	0.889
	PWY-6590: superpathway of Clostridium acetobutylicum acidogenic fermentation	0.054	0.638	0.931	-0.111	0.330
Propionate	ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis	0.191	0.092	0.384	0.106	0.351
	HISDEG-PWY: L-histidine degradation I	-0.056	0.621	0.996	-0.081	0.476
	HOMOSER-METSYN-PWY: L-methionine biosynthesis I	-0.004	0.972	0.951	-0.049	0.669
	HSERMETANA-PWY: L-methionine biosynthesis III	0.042	0.711	0.747	-0.059	0.604
	MET-SAM-PWY: superpathway of S-adenosyl-L-methionine biosynthesis	-0.034	0.768	0.919	-0.081	0.475
	METSYN-PWY: L-homoserine and L-methionine biosynthesis	-0.021	0.857	0.976	-0.062	0.585
	P108-PWY: pyruvate fermentation to propanoate I	-0.088	0.442	0.661	0.065	0.570
	P4-PWY: superpathway of L-lysine, L-threonine and L-methionine biosynthesis I	0.077	0.501	0.730	-0.062	0.589

	PWY-5028: L-histidine degradation II	0.127	0.267	0.689	-0.108	0.344
	PWY-5100: pyruvate fermentation to acetate and lactate II	-0.329	**0.003**	0.979	-0.122	0.283
	PWY-5345: superpathway of L-methionine biosynthesis (by sulfhydrylation)	-0.003	0.977	0.891	-0.025	0.829
	PWY-5347: superpathway of L-methionine biosynthesis (transsulfuration)	-0.014	0.899	0.979	-0.061	0.594
	PWY-5505: L-glutamate and L-glutamine biosynthesis	-0.202	**0.075**	0.747	-0.076	0.505
	PWY-6151: S-adenosyl-L-methionine cycle I	-0.183	0.107	0.976	-0.107	0.349
	PWY-6628: superpathway of L-phenylalanine biosynthesis	0.223	0.048	0.591	0.115	0.313
	PWY0-1061: superpathway of L-alanine biosynthesis	-0.042	0.711	0.591	-0.103	0.364
	PWY0-781: aspartate superpathway	0.077	0.502	0.730	-0.055	0.632
	SER-GLYSYN-PWY: superpathway of L-serine and glycine biosynthesis I	0.322	**0.004**	0.979	0.118	0.302
	THRESYN-PWY: superpathway of L-threonine biosynthesis	-0.032	0.782	0.839	-0.109	0.338

Correlation between mWGS key MetaCyc pathways profiles with its paired butyrate and propionate metabolites levels and replication of correlations between IPCO inferred profiles with butyrate and propionate levels. Significance for the mWGS profile and IPCO profiles is determined by nominal p-value ≤ 0.1 or p-adjusted ≤ 0.1 and highlighted in red and bold. mWGS profiles were not significantly correlated with butyrate and only three pathways with propionate levels. Directionality and strength of correlation is determined by spearman correlation estimate.

Chapter V

Gut microbiota dynamics in patients with Multiple Sclerosis

Jillian R-M Brown[#1,2], Mrinmoy Das[#1,2], Fergus Shanahan [1,3], Brian Sweeney [3], Ian B. Jeffery[1,2], Paul W. O'Toole[1,2]

# Contributed equally

(Manuscript in preparation)

1. APC Microbiome Ireland, University College Cork, Ireland

2. School of Microbiology, University College Cork, Ireland

3. Department of Medicine, Cork University Hospital, Ireland

Contents

Chapter 5

Chapter 5 Gut microbiota dynamics in patients with Multiple Sclerosis

5.1 Abstract

Background Multiple sclerosis (MS) is an immune-associated disorder of the neurons characterised by demyelination and inflammation. Microbiota alterations have been identified as a risk factor in various inflammatory conditions. However, age-related microbiota alterations related to biological and premature ageing have not been previously considered as a confounder.

Aim To investigate the gut microbiota compositional changes associated with the different phenotypes of MS.

Method This study consists of 32 MS patients stratified into 21 relapse-remitting MS (RRMS), 8 secondary progressive MS (SPMS) and 3 primary progressive MS (PPMS) phenotypes. These samples are compared against 9 young healthy and 96 elderly healthy samples. Whole genome shotgun (mWGS) sequencing of the faecal samples was carried out and analysed to identify taxonomic and functional changes associated with MS phenotypes.

Results We identified significant separation in the β-diversity associated with the group stratification. Increased abundances of taxa including *Clostridium* species, *Eggerthella* and *Escherichia coli* were observed in MS patients, whereas *Lachnospiracea, Eubacterium* and *Ruminococcus* that are previously associated with high fibre diet and improved health were less abundant. Functional profiles associated with fermentation and secondary metabolites were reduced in MS patients.

Conclusion We identified taxonomic and functional alterations associated with MS that are distinct from age-related phenotypes. These changes reflect the depletion and loss of health associated taxa.

Keywords

Multiple sclerosis, gut microbiota, inflammation, ageing

5.2 Introduction

Multiple sclerosis (MS) is a neurodegenerative disorder of the central nervous system (CNS) involving autoimmune responses and premature cellular aging (Nicaise et al., 2019). This results in lesions due to demyelination in the nerve cells in optic nerves, white and grey matter (Filippi et al., 2018). The aetiology of the disease is multi-factorial and includes environmental stressors but not ageing. MS is more prevalent in western countries with European ancestry compared to Asian, African or Native American population (Rosati, 2001) and is commonly reported in adults aged 20-40 years (Ghasemi et al., 2017).

MS is classified into different subtypes depending on the onset, progression, and assessment of the disease. These subtypes include clinically isolated syndrome (CIS) defined by initial onset of the disease. Relapsing-remitting MS (RRMS) is characterised by reversible relapse of the lesions. Primary progressive MS (PPMS) is diagnosed with disease progression continuing from the beginning without relapses. Once MS progression results in permanent defects, it is considered as secondary progressive MS (SPMS) (Filippi et al., 2018). Polymorphism in various immune associated genes like HLA class I and II, and Interleukin (IL) genes have been identified as risk factors (Cotsapas and Mitrovic, 2018). Production of pro-inflammatory cytokines, influx of immune cells, ineffective functioning of the T-regulatory cells are reported to cause neuro-axonal damage (Filippi et al., 2018).

The gut microbiota is associated with various immune functions of the host that involve T-cells mediated immune homeostasis (Pandiyan et al., 2019). Experimental autoimmune encephalomyelitis (EAE) animal models are used for studying MS and they highlight T-cell mediated immune response resulting in demyelination of neurons (Berer et al., 2011;Lee et al., 2011). Germ-free EAE models

or antibiotics treatment of animal models were observed to have reduced disease activity and increased immune tolerance (Yokote et al., 2008;Lee et al., 2011). Alternatively, certain bacterial metabolites (short chain fatty acids) were shown to have protective activity against EAE (Haghikia et al., 2015). A reduction in disease severity is observed in animal models upon colonisation of certain species from *Lactococcus, Bifidobacterium* and *Prevotella* (Ezendam et al., 2008;Rezende et al., 2013;Mangalam et al., 2017).

Microbial alterations have been associated with EAE susceptibility in animal models and as well as in MS patients. In human studies, patients with MS were shown to have a reduced abundance of taxa associated with short-chain fatty acid production, microglia and T-regulatory cells interactions (Mirza and Mao-Draayer, 2017). Despite identification of taxonomic association with MS, a recent study observed that the predictive power of the taxonomic features was very low with high false negative rates across multiple methods (Bang et al., 2019).

In the current study, we hypothesised that the gut microbiota profiles might be distinct in MS patients with different unique and shared signature between subtypes, but that signature might be confounded by alterations associated with premature cellular aging. Therefore, the MS signature is compared against young healthy and an elderly healthy population to understand the changes in microbiota dynamics that is altered in the disease and associated subgroups.

5.3 Method

5.3.1 Sample collection

Recruitment of samples was carried out in Cork University Hospital, Cork. Subjects diagnosed with RRMS, PPMS and SPMS were recruited along with healthy volunteers. Written consent was obtained from all recruited subjects. Exclusion criteria were defined as antibiotic usage in 3 months prior sampling, history of alcohol abuse or participation in drug trials 1 month prior.

5.3.2 DNA Extraction

Genomic DNA was extracted from faecal samples (0.25 g) using the Repeat Bead Beating (RBB) method of Yu and Morrison (Yu and Morrison, 2004) with the following modifications. Sterile zirconia beads (0.5 g) collection comprising of one 3.0 mm bead, 0.1 g of 0.5 mm beads, and 0.3 g of 0.1 mm beads were used. Faecal samples were homogenised via bead beating for 90 seconds (Mini-Beadbeater™, BioSpec Products, Bartlesville, OK, USA), with the samples cooled on ice for 60 seconds before another 90 seconds bead beating. Pooled supernatants were incubated with 350 ml of 7.5 M ammonium acetate (Sigma). The extraction then proceeded as per the RBB extraction protocol. Genomic DNA was visualised on 1% agarose gel and quantified using the Nanodrop 1000 (Thermo Scientific, Ireland). Extracted genomic DNA was stored at -20°C until sent for metagenomic whole genome shotgun (mWGS) sequencing in HiSeq 2000 sequencer at Beijing Genomics Institute-BGI (China).

5.3.3 Bioinformatics analysis

Quality filtering and trimming of mWGS sequence data was carried out as described in HMP project (Lloyd-Price et al., 2017). Human reads were removed using BMTagger (v.3.1) by indexing it against human reference genome (hg19). Low quality sequences and reads shorter than 60 bases were removed using trimBWAstyle.pl. Taxonomic and functional profiles were obtained using MetaPhlAn2 (Truong et al., 2015) and HUMAnN2 (Franzosa et al., 2018) respectively with default parameters. UniRef90 database was used in HUMAnN2 pipeline. Counts per million (CPM) normalisation was carried out on pathway abundances output. Community samples from the ELDERMET cohort were included into the study for comparison against an aging population (Claesson et al., 2012). Species-level taxonomic dataset and pathway dataset stratified with taxonomic information was used for all analysis.

5.3.4 Statistical Analysis

All analyses were carried out in R (v.3.3.1) (Team, 2016). Using a threshold of P-adjusted ≤ 0.05 (Benjamini-Hochberg method), statistical significance was defined. Alpha diversity was measured using observed species index. Alterations in global microbiota profile was investigated using redundancy analysis (RDA) from vegan library (v. 2.4.3) (Oksanen et al., 2017) on log10-transformed data modelled against the group stratification and visualised using s.class functionality from the ade4 (v. 1.7.6) library (Dray and Dufour, 2007). A minimum value of 1e-05 was added to the taxonomic data before log transformation. Significance of the RDA model and effect size was calculated using analysis of variance (ANOVA) with 1000 permutations.

Species level taxonomic data and pathways abundance profiles were filtered to retain only those features that were present in at least 50% samples in at least one

group. Differentially abundant taxa and pathways were determined using Kruskal-Wallis test and Dunn's test *post-hoc* analysis. P-values from Kruskal-Wallis were adjusted and pairwise analysis was carried out only on those features that passed the p-adjusted threshold from Kruskal-Wallis test.

Binary dataset was generated from the pairwise comparison carried out in the taxonomic and functional datasets. Taxonomic and functional features that were significant (P-adjusted ≤ 0.05) were defined as 1 and non-significant features were denoted as 0. Binary distance for the features was calculated and clustered with Ward's linkage (ward.D2 method; R statistical software). The binary dataset and clustered dendrogram was visualised as a heatmap from gplots (v. 3.0.1) (Warnes et al., 2019) to identify clusters associated with different group stratification. Identification of clusters was carried out by visual observation. Enrichment/depletion of significant pathways in each cluster was carried out using Fisher's exact test by comparing the pathway representation within the cluster versus across the whole dataset excluding the cluster. Nominal P-value (≤ 0.05) was used to define statistical significance from the Fisher's exact test. Bar plots were created using ggplot2 (v. 2.2.1) (Wickham, 2009).

5.4 Results

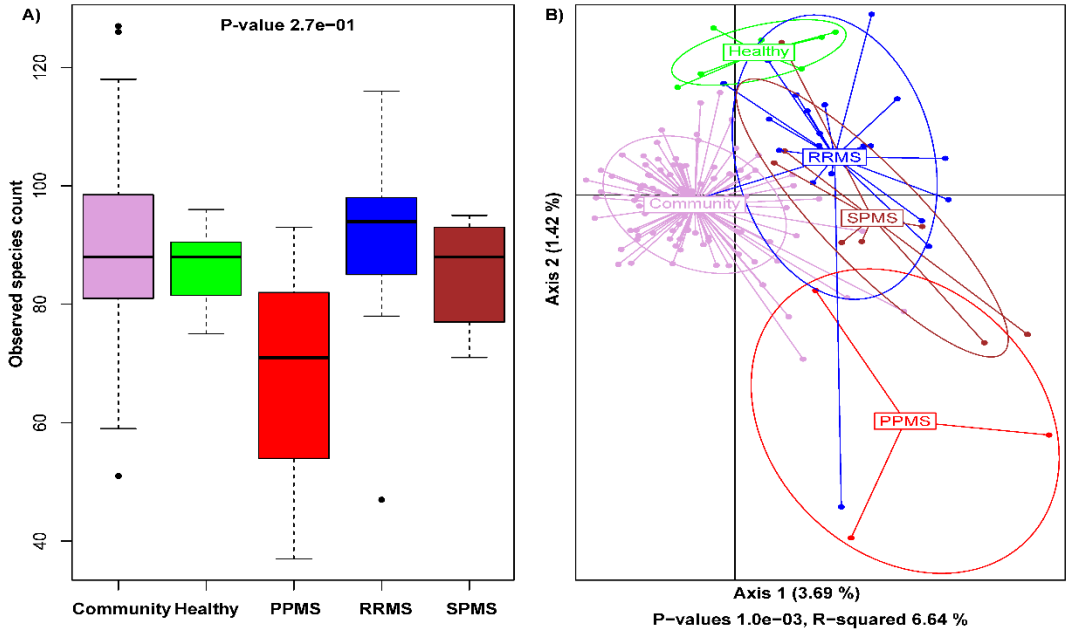
5.4.1 Sample characteristics

A total of 32 MS samples were considered for the study. The controls samples comprised of 7 young healthy and 96 ELDERMET community samples. Overall, the study included 96 community dwelling elderly, 7 healthy, 21 RRMS, 8 SPMS and 3 PPMS samples. From the MS samples plus 7 young healthy samples, a total of 174 gigabases of quality filtered sequencing data with 25,050,928±1,228,109 reads per sample was obtained.

5.4.2 Gut microbiota profiles of the cohort

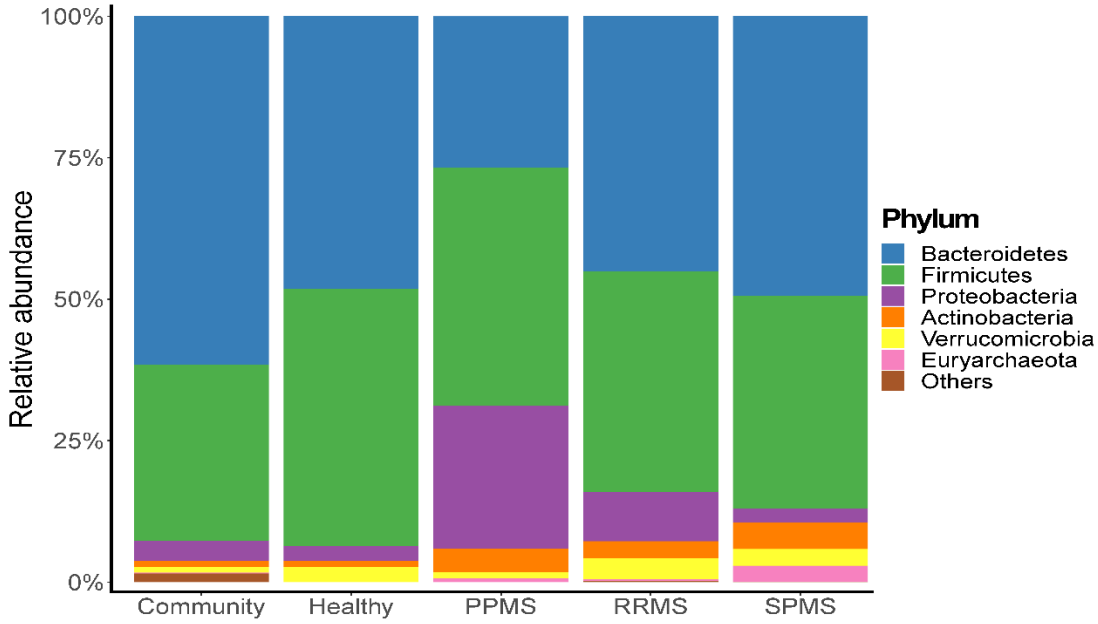
From the mWGS sequencing run, including community samples, a total of 399 species belonging to 66 different families were profiled. The α-diversity defined by number of observed species showed no significant differences between groups **(Figure 1A, Table 1)**. The RDA analysis on the global microbiota composition identified significant differences (p-value ≤ 0.05) in the microbiota structure based on group stratification with an effect size of 6.64% **(Figure 1B)**. The separation between different subtypes of MS and young healthy was observed along the Y-axis explaining 1.42 % of variance whereas the X-axis identified separation between the community samples and MS subtypes with 3.69 % of variance. The SPMS and RRMS did not show prominent separation between themselves on the 1st two redundancy axes. The young healthy and the elderly community samples also separated along the Y-axis. Based on the relative abundance at Phylum level, Bacteroides and Firmicutes dominated, followed by Proteobacteria and Actinobacteria across all samples **(Figure 2)**.

Figure 1 Taxonomic diversity within and between Groups



A) Represents the observed species count within each group. Non-significant (P-value ≤ 0.05) difference is observed between the groups. **B)** Ordination based on redundancy analysis on the log10 transformed taxonomic species-level dataset. Significant separation is observed between MS phenotypes on the Y-axis and with the community elderly samples on the X-axis.

Figure 2 Mean relative abundance at Phylum level



The average relative abundance of each phyla observed across the whole dataset stratified by group status.

Table 1 Observed taxonomic and functional diversity

Groups	Observed species	Observed gene count
Community	89.1 ± 14.2	286589.8 ± 56489.9
Healthy	86.1 ± 7.9	321115.3 ± 23505.4
PPMS	67 ± 28.2	205051 ± 163008.9
RRMS	91.1 ± 13.8	313949.3 ± 45406.6
SPMS	85.3 ± 9.4	274191 ± 40564.2

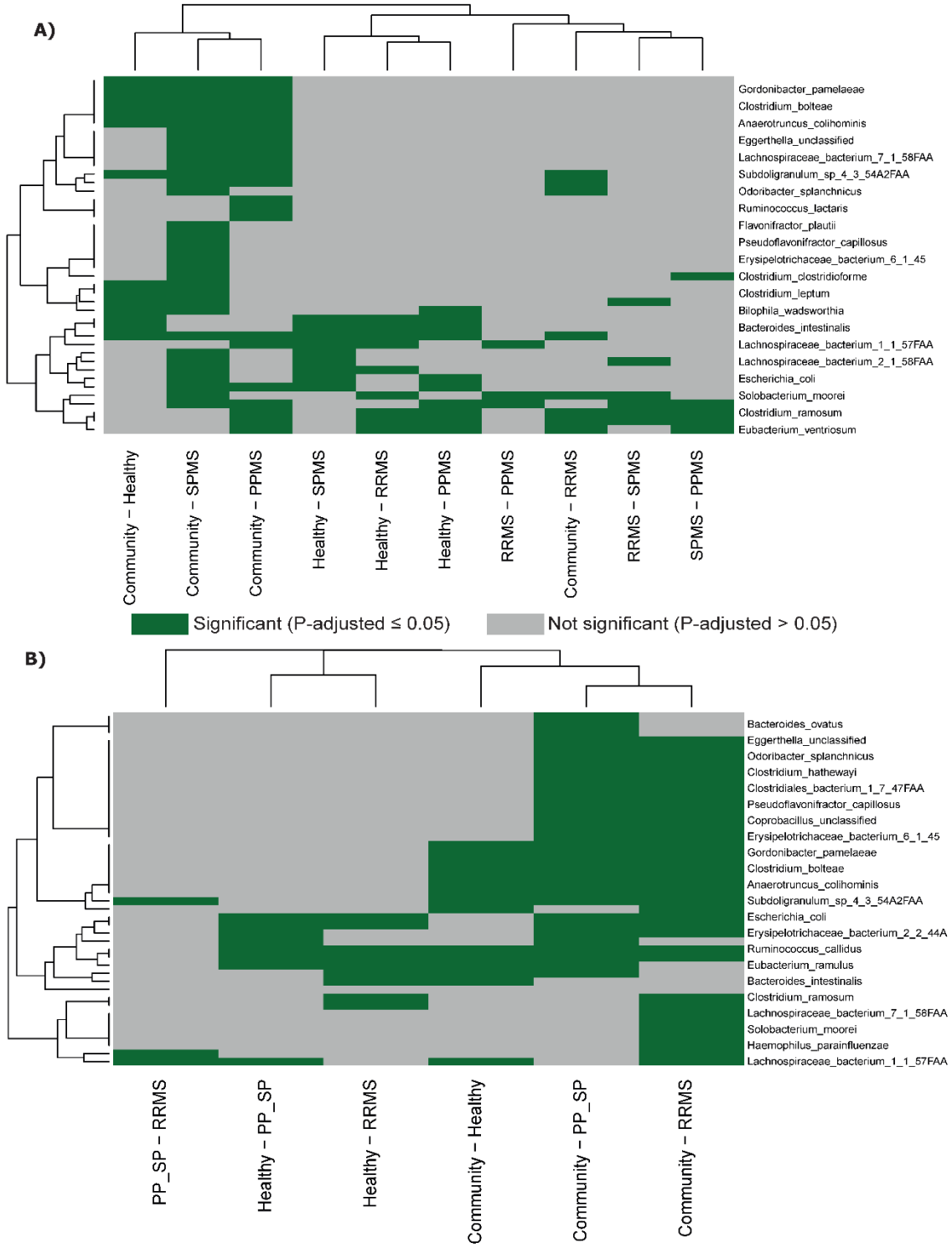
The values reported are the average values ± standard deviation for each group.

5.4.3 Taxonomic abundance was significantly altered in MS groups

The identification of significantly differentially (P-adjusted ≤ 0.05) abundant species was carried out on a filtered taxonomic dataset containing abundance information for 117 species. Based on Kruskal-Wallis test, we identified 42 taxa in total to be significantly different in at least one group. These significant taxa were further investigated using pairwise comparisons, as represented in **Figure 3A**. PPMS and SPMS partly shared similar significance results, hence despite being defined as different MS subtypes medically, were merged together as an artificial group represented by PP_SP. Based on this new grouping we identified 44 species to be significantly different in at least one group. *Post-hoc* analysis revealed significant taxonomic signature associated with the groups **Figure 3B**. The mean and standard deviation of the relative abundances of the significant taxa per groups are represented in **Table 2**. In summary, the MS phenotype groups identified different *Clostridium* species to be significantly different from elderly samples. Most species were observed to be more abundant in MS compared to elderly samples. In the RRMS groups, *Clostridium* and *Eggerthella* species were more abundant whereas *Lachnospiracea, Ruminococcus* and *Hemophilus parainfluenzae* were depleted in comparison to elderly samples. Similar trends were observed in PPMS and SPMS merged group, which also included reduced abundance of *Bacteroides ovatus*. Compared to young healthy

samples, very few species were significantly different in different MS phenotypes. This included increased abundance of *Escherichia coli, Clostridium bolteae* and decreased abundance of *Ruminococcus callidus, Lachnospiraceae bacterium* and *Eubacterium ramulus* in MS groups. The elderly samples were distinguished from the healthy with reduced abundances of *Clostridium bolteae, Ruminococcus callidus, Bacteroides intestinalis, Eubacterium ramulus* and *Lachnospiraceae.*

Figure 3 Taxonomic alteration associated with aging and MS stratification.



A) Heatmap of the significantly differently abundant species which highlights the significant taxa across pairwise comparisons. Clustering is based on binary distance using Ward's linkage. **B)** Heatmap representing significantly different species based on merged PPMS and SPMS groups (PP_SP). Clustering is based on binary distance and Ward's linkage.

Table 2 Taxonomic abundance across the control and MS stratifications.

Significant species	Community	Healthy	PP_SP	RRMS
Anaerostipes hadrus	0.06 ± 0.16	0.23 ± 0.32	0 ± 0.01	0.01 ± 0.02
Anaerotruncus colihominis	0 ± 0.01	0.02 ± 0.03	0.05 ± 0.09	0.08 ± 0.13
Bacteroides intestinalis	0.66 ± 2.17	2.23 ± 4.26	0.92 ± 3.05	1.14 ± 2.98
Bacteroides ovatus	2.62 ± 3.73	1.18 ± 1.36	0.45 ± 0.56	1.24 ± 1.9
Bifidobacterium longum	0.21 ± 0.49	0.3 ± 0.36	2.44 ± 5.02	1.49 ± 3.95
Bilophila unclassified	0.18 ± 0.24	0.53 ± 0.3	0.35 ± 0.36	0.79 ± 0.84
Bilophila wadsworthia	0.02 ± 0.03	0.09 ± 0.05	0.05 ± 0.06	0.11 ± 0.15
Clostridiales bacterium_1_7_47FAA	0 ± 0	0.01 ± 0.01	0.01 ± 0.03	0.01 ± 0.02
Clostridium asparagiforme	0 ± 0.01	0.02 ± 0.02	0.04 ± 0.06	0.06 ± 0.08
Clostridium bolteae	0.02 ± 0.06	0.11 ± 0.23	0.58 ± 0.96	0.25 ± 0.78
Clostridium citroniae	0 ± 0.01	0.02 ± 0.02	0.02 ± 0.03	0.03 ± 0.04
Clostridium clostridioforme	0 ± 0	0.01 ± 0.02	0.02 ± 0.06	0.19 ± 0.51
Clostridium hathewayi	0.07 ± 0.68	0 ± 0.01	0.22 ± 0.28	0.03 ± 0.03
Clostridium leptum	0.02 ± 0.05	0.11 ± 0.14	0.04 ± 0.05	0.06 ± 0.07
Clostridium ramosum	0 ± 0.01	0 ± 0	0.03 ± 0.04	0 ± 0.01
Clostridium symbiosum	0 ± 0.01	0.01 ± 0.02	0.23 ± 0.39	0.06 ± 0.09
Coprobacillus unclassified	0.01 ± 0.08	0 ± 0	0.06 ± 0.09	0.01 ± 0.02
Eggerthella lenta	0 ± 0.01	0 ± 0	0.07 ± 0.22	0.01 ± 0.02
Eggerthella unclassified	0.01 ± 0.02	0.02 ± 0.03	0.88 ± 2.54	0.13 ± 0.3
Erysipelotrichaceae bacterium_2_2_44A	0 ± 0.02	0 ± 0	0.03 ± 0.04	0.01 ± 0.02
Erysipelotrichaceae bacterium_21_3	0 ± 0.01	0 ± 0	0.01 ± 0.01	0.01 ± 0.02
Erysipelotrichaceae bacterium_6_1_45	0 ± 0.01	0 ± 0	0.01 ± 0.01	0.01 ± 0.01
Escherichia coli	1.35 ± 5.03	0.05 ± 0.1	0.52 ± 0.92	2.32 ± 3.94
Eubacterium ramulus	0.06 ± 0.08	0.52 ± 0.7	0.06 ± 0.11	0.09 ± 0.16
Eubacterium ventriosum	0.12 ± 0.18	0.19 ± 0.3	0.01 ± 0.02	0.07 ± 0.1
Flavonifractor plautii	0.01 ± 0.01	0.01 ± 0.02	0.03 ± 0.08	0.04 ± 0.08
Gordonibacter pamelaeae	0 ± 0	0.01 ± 0.02	0.02 ± 0.04	0.02 ± 0.03
Haemophilus parainfluenzae	0.59 ± 1.54	0.11 ± 0.2	0.03 ± 0.06	0.03 ± 0.06
Holdemania filiformis	0.01 ± 0.03	0.01 ± 0.01	0.03 ± 0.04	0.04 ± 0.04
Holdemania unclassified	0 ± 0.01	0.03 ± 0.05	0 ± 0.01	0.02 ± 0.03
Lachnospiraceae bacterium_1_1_57FAA	0.07 ± 0.21	0.16 ± 0.22	2.04 ± 5.87	0.1 ± 0.24
Lachnospiraceae bacterium_3_1_57FAA_CT1	0.02 ± 0.05	0 ± 0	0.51 ± 1.25	0.03 ± 0.08
Lachnospiraceae bacterium_5_1_63FAA	0.04 ± 0.11	0.05 ± 0.05	0.01 ± 0.03	0.02 ± 0.03
Lachnospiraceae bacterium_7_1_58FAA	0.03 ± 0.09	0.04 ± 0.06	0.13 ± 0.16	0.08 ± 0.09
Leuconostoc citreum	0.01 ± 0.01	0 ± 0	0 ± 0	0 ± 0
Odoribacter splanchnicus	1.71 ± 1.23	0.94 ± 0.64	0.75 ± 0.94	0.88 ± 0.74
Oscillibacter unclassified	0.19 ± 0.23	0.39 ± 0.41	0.76 ± 0.66	0.85 ± 0.77
Pseudoflavonifractor capillosus	0 ± 0.01	0 ± 0	0.01 ± 0.02	0.01 ± 0.02
Ruminococcus callidus	0.12 ± 0.27	0.56 ± 0.88	0.01 ± 0.02	0.01 ± 0.03
Ruminococcus lactaris	0.27 ± 0.5	0.35 ± 0.47	0 ± 0	0.12 ± 0.23
Solobacterium moorei	0 ± 0	0 ± 0	0.02 ± 0.07	0 ± 0
Subdoligranulum sp_4_3_54A2FAA	0.01 ± 0.07	0.19 ± 0.21	1.04 ± 2.59	0.47 ± 1.2
Subdoligranulum unclassified	5.41 ± 5.5	12.82 ± 9.88	11.41 ± 10.62	7.59 ± 5.98
Veillonella parvula	0.12 ± 0.33	0.01 ± 0.01	0.01 ± 0.02	0.04 ± 0.07

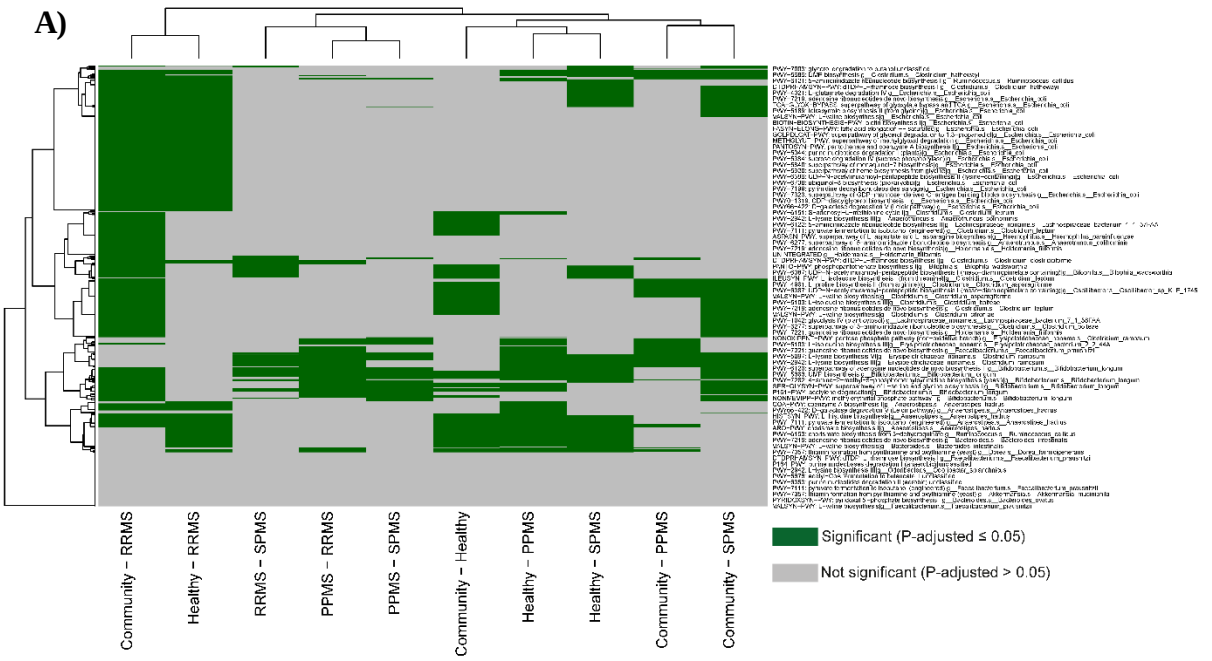
The values reported are the average relative abundances ± standard deviation.

5.4.4 Differential functionality associated with control and MS stratifications

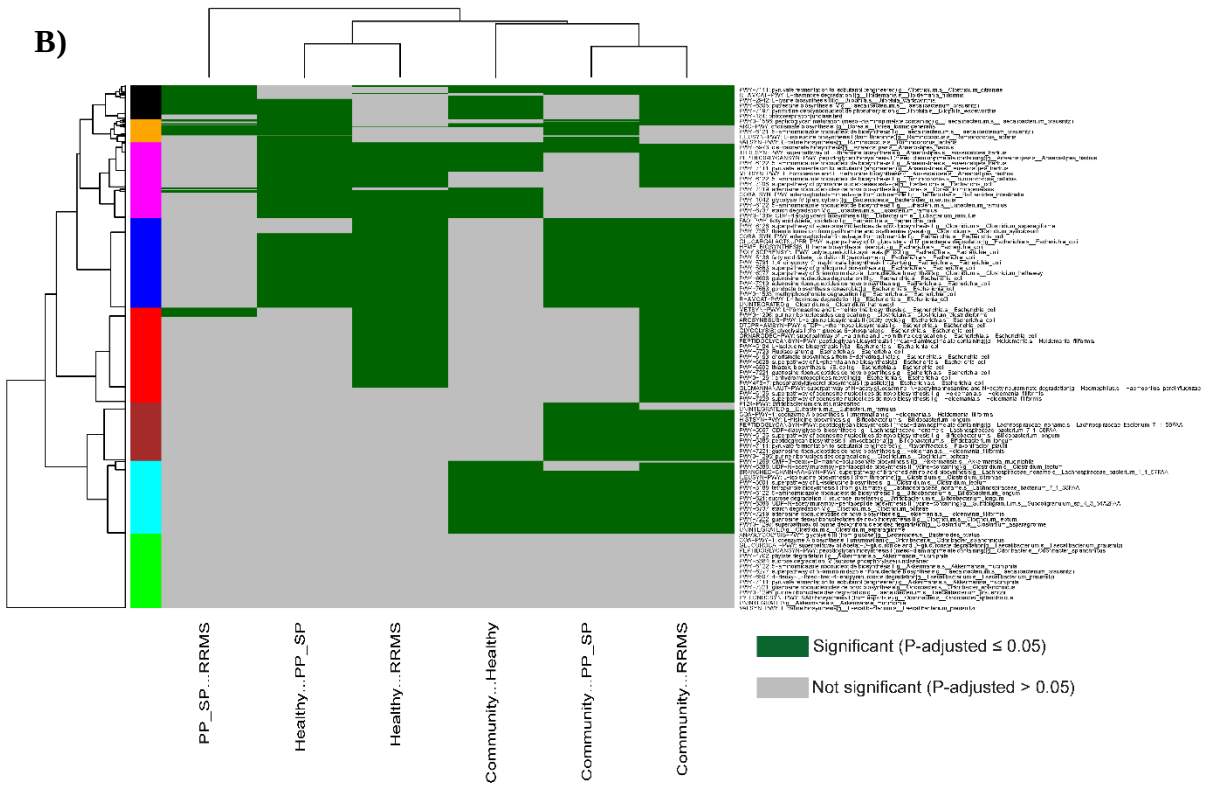
The functional diversity defined by the number of genes identified is described in **Table 1**. There was a trend towards significance in the difference in gene counts between the groups (P-value: 0.1). The identified UniRef genefamilies corresponded to 439 MetaCyc pathways belonging to 280 species. After filtering, 330 pathways were retained that belonged to 84 species. Differential abundance analysis with Kruskal-Wallis test identified 275 pathways from 38 species as significantly (P-adjusted ≤ 0.05) different in at least one group. *Post-hoc* analysis from the significant outcome of Kruskal-Wallis test identified 198 pathways that showed significance in pairwise comparisons. Binary clustering of the significant (P-adjusted ≤ 0.05) pathways identified unique species-specific pathways signature associated with different groups amongst which PPMS and SPMS showed shared signature **(Figure 4A)**.

Merging the two groups: PPMS and SPMS as PP_SP and reanalysing the pathway abundance identified 283 pathways from 36 species as significantly different amongst the groups. Binary clustering on the regrouped data identified prominent clusters **(Figure 4B)**. All MS phenotypes cluster (cluster blue) was associated with 132 pathways from 8 species. RRMS cluster (cluster red) contains 144 pathways from 8 species. The merged PPMS and SPMS cluster (cluster orange) has 32 pathways from 9 species. The healthy cluster (cluster magenta) is associated with 82 pathways from 9 species. The elderly cluster (cluster cyan) has 59 pathways from 14 species. The elderly vs MS cluster (cluster brown) is associated with 65 pathways from 16 species. Cluster green contains 82 pathways that were identified in the multigroup analysis but showed no pairwise significance.

Figure 4 Functional species-specific pathways are differentially altered across control and MS stratifications



Heatmap representing significantly (P-adjusted ≤ 0.05) different pathways with species stratification based on clinical groups. Clustering is based on binary distance using ward.D2.



Heatmap representing significantly (P-adjusted ≤ 0.05) different pathways with species stratification based on merged PPMS and SPMS groups (PP_SP). Clustering is based on binary distance using ward.D2.

Each cluster was further evaluated for enriched/depleted pathways. Identification of significantly (P-value ≤ 0.05) enriched/depleted pathways from each cluster is detailed in **Table 3**. Interestingly cluster blue which is associated with all MS phenotype did not identify any significantly enriched/depleted pathways. In RRMS associated cluster, only two pathways were significantly different. The two pathways "PWY-5083: NAD/NADH phosphorylation and dephosphorylation" and "PWY-6892: thiazole biosynthesis I (E. coli)" both belongs to superclass of cofactors, electron carrier and vitamin biosynthesis. These two pathways were more abundant in RRMS compared to healthy. In the elderly associated cluster, the pathways identified were primarily associated with different *Clostridium* species. Three of the pathways were nucleic acid biosynthesis, L-profile biosynthesis, and nitrate reduction. All of these pathways have decreased abundance in elderly samples compared to young healthy and MS samples. The brown cluster that is associated with difference between elderly and MS phenotypes has pathways associated with secondary metabolite production (methylerythritol phosphate pathway I) and fermentation process (Bifidobacterium shunt) which are more abundant in the elderly samples. The healthy associated cluster retained maximum number of enriched/depleted pathways. The mixed SPMS and PPMS cluster identified a significantly different archaeal pathway (Flavin biosynthesis) that was more abundant in the MS samples compared to community and healthy samples. Valine biosynthesis and peptidoglycan biosynthesis pathways were less abundant in SPMS-PPMS merged group compared to healthy and community samples.

Table 3 List of pathways that are significantly enriched/depleted in each cluster

Clusters	Pathways	Species	Community	Healthy	PP_SP	RRMS	P-value
red	PWY-5083: NAD/NADH phosphorylation and dephosphorylation	*Escherichia coli, Haemophilus parainfluenzae*	6.99 ± 14.98	1.05 ± 1.4	5.06 ± 8.76	12.27 ± 20.74	3.0E-03
	PWY-6892: thiazole biosynthesis I (E. coli)	*Escherichia coli*, unclassified	8.51 ± 11.8	3.29 ± 3.27	5.31 ± 7.88	7.95 ± 8.99	2.7E-02
black	PWY-181: photorespiration	unclassified	0.47 ± 1.22	1.73 ± 1.73	0.88 ± 1.71	1.29 ± 2.67	1.9E-02
	PWY-6737: starch degradation V	*Clostridium citroniae, Clostridium leptum, Clostridium symbiosum*	0.09 ± 0.27	0.91 ± 1.05	0.92 ± 1.68	0.87 ± 1.03	3.4E-02
cyan	PWY-4981: L-proline biosynthesis II (from arginine)	*Clostridium asparagiforme, Clostridium citroniae*	0 ± 0.03	0.24 ± 0.24	0.6 ± 0.91	0.33 ± 0.31	4.2E-02
	PWY-7219: adenosine ribonucleotides de novo biosynthesis	*Anaerotruncus colihominis, Bifidobacterium longum, Clostridium asparagiforme, Clostridium leptum, Holdemania filiformis, Lachnospiraceae bacterium_1_1_57FAA, Subdoligranulum sp_4_3_54A2FAA*	1.08 ± 1.67	6.4 ± 5.82	29.33 ± 50.16	13.38 ± 26.99	4.4E-02
	PWY-7220: adenosine deoxyribonucleotides de novo biosynthesis II	*Clostridium asparagiforme, Clostridium leptum, Subdoligranulum sp_4_3_54A2FAA*	0.09 ± 0.29	1.53 ± 1.36	3.66 ± 6.71	1.52 ± 2.05	1.5E-02
	PWY-7222: guanosine deoxyribonucleotides de novo biosynthesis II	*Clostridium asparagiforme, Clostridium leptum, Subdoligranulum sp_4_3_54A2FAA*	0.09 ± 0.29	1.53 ± 1.36	3.66 ± 6.71	1.52 ± 2.05	1.5E-02
	PWY490-3: nitrate reduction VI (assimilatory)	unclassified	0.3 ± 0.63	1.55 ± 1.5	0.27 ± 0.46	1.21 ± 1.95	4.2E-02
magenta	CRNFORCAT-PWY: creatinine degradation I	unclassified	1.93 ± 2.42	5.39 ± 3.34	2.13 ± 2.75	3.33 ± 3.45	4.4E-02
	HCAMHPDEG-PWY: 3-phenylpropanoate and 3-(3-hydroxyphenyl) propanoate degradation to 2-oxopent-4-enoate	*Escherichia coli*	1.79 ± 5.53	0 ± 0	1.88 ± 3.46	6.88 ± 14.89	4.4E-02
	PWY-5022: 4-aminobutanoate degradation V	unclassified	0.73 ± 0.98	1.72 ± 0.63	0.63 ± 0.65	0.65 ± 0.93	4.4E-02

	PWY-5918: superpathay of heme biosynthesis from glutamate	*Escherichia coli*	2.66 ± 8.48	0.15 ± 0.4	3.21 ± 5.9	7.58 ± 12.44	4.4E-02
	PWY-6690: cinnamate and 3-hydroxycinnamate degradation to 2-oxopent-4-enoate	*Escherichia coli*	1.79 ± 5.53	0 ± 0	1.88 ± 3.46	6.88 ± 14.89	4.4E-02
	PWY-7269: NAD/NADP-NADH/NADPH mitochondrial interconversion (yeast)	*Escherichia coli*	3.21 ± 10.16	0.21 ± 0.56	3.64 ± 6.6	8.61 ± 14.78	4.4E-02
	PWY0-1277: 3-phenylpropanoate and 3-(3-hydroxyphenyl) propanoate degradation	*Escherichia coli*	3.05 ± 10.08	0 ± 0	3.53 ± 6.83	10.57 ± 20.94	4.4E-02
	PWY0-1298: superpathway of pyrimidine deoxyribonucleosides degradation	*Anaerostipes hadrus, Escherichia coli*	4.2 ± 10.78	2.86 ± 3.44	3.87 ± 6.88	9.61 ± 15.69	2.5E-02
brown	NONMEVIPP-PWY: methylerythritol phosphate pathway I	*Alistipes indistinctus, Bifidobacterium longum, Odoribacter splanchnicus*	5.92 ± 3.98	4.54 ± 2.5	5.62 ± 3.04	4.7 ± 4.4	4.9E-02
	P124-PWY: Bifidobacterium shunt	unclassified	1.22 ± 1.25	0.3 ± 0.8	0.62 ± 1.15	0.54 ± 1.17	3.3E-02
orange	PWY-6167: flavin biosynthesis II (archaea)	*Methanobrevibacter smithii*	0.28 ± 0.73	0 ± 0	1.32 ± 2.1	5.17 ± 20.4	1.3E-02
	PWY0-1586: peptidoglycan maturation (meso-diaminopimelate containing)	*Bacteroides ovatus, Faecalibacterium prausnitzii*	33.13 ± 30.8	51.89 ± 43.27	22.71 ± 28.49	22.41 ± 23.22	1.4E-02
	VALSYN-PWY: L-valine biosynthesis	*Clostridium leptum, Dorea formicigenerans, Ruminococcus lactaris*	1.78 ± 2.26	4.3 ± 2.87	0.68 ± 0.63	1.59 ± 1.59	4.9E-02

The values reported are the average abundance values ± standard deviation for each group. Significance is defined by nominal P-value ≤ 0.05. The species column reports the species for which each significant pathway was stratified with.

5.5 Discussion

The importance of gut microbiota in immune homeostasis and immune-mediated disorders is well established. Amongst all the MS phenotypes, RRMS is the most commonly identified form (Doshi and Chataway, 2016). The environmental influence defined by microbiota and its metabolites in MS is established using EAE animal models (Yokote et al., 2008;Lee et al., 2011). This study explores the taxonomic signature and potential functionality associated with different MS phenotypes and its deviation from a healthy ageing microbiota signature. It identified dissimilarity in global taxonomic composition between different MS phenotypes compared to both young and elderly healthy samples. This was reflected through differential abundances of both taxonomic and functional markers which are associated with the different MS phenotypes.

Diversity plays a key role in maintenance of the immune system. Studies have shown that presence or depletion of certain taxa can trigger or supress inflammation (Forbes et al., 2016). In the current study, we observed that the dissimilarity between the samples was distinct despite small effect size.

Further exploration of the taxa associated with the different groups have identified increased abundances of certain *Clostridium* species in the MS samples. Various studies have identified different species of this genus to be associated with pro- or anti-inflammatory properties depending on the species (Lopetuso et al., 2013). Similarly, *Eggerthella* is also reported to be higher in cases where T-cell functionality is partly compromised (Jin et al., 2015). Certain species from *Lachnospiracea*, *Eubacterium* and *Ruminococcus* are associated with health beneficial activities like T-regulatory cell modulation and are observed in healthy controls across different studies (Kang et al., 2010). These species are known producers of short-chain fatty acids,

which is linked with high-fibre diet and improved health. The altered abundances of these species in the current study could be related to potentially improper immune modulation in the gut. Similarly, polysaccharides from different species of *Bacteroides* have been identified to have different effect in immune modulation (Wexler, 2007). In this regards, *Bacteroides ovatus* is identified with increased levels of immunoglobulin A (IgA) and IgG antibodies in inflammatory bowel disease patients however, their exact role in pathogenesis is unclear (Saitoh et al., 2002). This species interestingly is reduced in our study in the MS phenotype groups. Prevalence of *Escherichia coli* and *Clostridium bolteae* in the MS samples could be associated with infection in MS (Kaper et al., 2004;Song et al., 2004). The young healthy and elderly samples also showed a difference in abundances of SCFA producers and *Bacteroides intestinalis* (Shortt et al., 2018). These changes may be associated with ageing, dietary habits, medications, and lifestyle changes.

The potential functionality of the microbiota in this study have highlighted different functional clusters associated with different groups. The MS associated clusters that showed difference against healthy young or elderly samples contains superclasses of pathways associated with secondary metabolism, fermentation, and cofactors biosynthesis. The contributing taxa to these functionalities are also observed to be the significantly different. Despite uncovering different functionality, their role in the pathogenesis of MS remains unclear.

This study undertook robust and rigorous statistical measures to investigate the microbiota from the MS cohort. Despite uncovering microbiota alterations, which are independently established in different studies, the study is limited by the sample size representation per group and a lack of clinical data. The merging of PPMS and SPMS due to statistically significance and similar signature would require further

investigation into the similarity and dissimilarity of the disease onset in these two sub phenotypes accounting for microbiota profiles. The functional potential identified in the study is not a representation of the actual activity of the microorganism and must be carefully considered. In conclusion, we report taxonomic and functional alterations that are associated with differences in the abundances of different species. These differences may be linked with immune modulation but appear to be independent of any age-related microbiota alterations associated with premature ageing. Further exploration may identify possible links with disease onset and progression of MS. These taxa may also serve as diagnostic measure to identify disease risk.

5.5 Acknowledgements

Acknowledgements

We kindly acknowledge the academic support infrastructure of the APC Microbiome Ireland, University College Cork, Cork.

Competing Interests

None to declare

Funding

The authors are supported in part by Science Foundation Ireland (SFI) in the form of a research centre grant (SFI/12/RC/2273). IBJ and MD are funded by an SFI Starting Investigator Research Grant (SIRG) (13/SIRG/2128).

Ethics approval

Clinical Research Ethics Committee of the Cork Teaching Hospitals (CREC)

Contributors

JRMB contributed to the concept, design of the work, acquisition of data, extraction, sequencing, and interpretation of the work, scripting and revising the draft. MD carried out the bioinformatics analysis, compilation and interpretation of the work, scripting and revising the draft. BS is a clinical liaison office who recruited the patients, contributed to the design of the work, acquisition of data, and reviewed the manuscript before submission. PWOT contributed to the concept and design of the work, interpretation of the work, scripting and revising the draft. FS contributed to the concept and design of the work and revising the draft. IBJ contributed to the concept and design of the work, interpretation of the work, scripting and revising the draft.

5.6 References

1. Bang, S., Yoo, D., Kim, S.J., Jhang, S., Cho, S., and Kim, H. (2019). Establishment and evaluation of prediction model for multiple disease classification based on gut microbial data. *Sci Rep* 9**,** 10189.
2. Berer, K., Mues, M., Koutrolos, M., Rasbi, Z.A., Boziki, M., Johner, C., Wekerle, H., and Krishnamoorthy, G. (2011). Commensal microbiota and myelin autoantigen cooperate to trigger autoimmune demyelination. *Nature* 479**,** 538-541.
3. Claesson, M.J., Jeffery, I.B., Conde, S., Power, S.E., O'connor, E.M., Cusack, S., Harris, H.M., Coakley, M., Lakshminarayanan, B., O'sullivan, O., Fitzgerald, G.F., Deane, J., O'connor, M., Harnedy, N., O'connor, K., O'mahony, D., Van Sinderen, D., Wallace, M., Brennan, L., Stanton, C., Marchesi, J.R., Fitzgerald, A.P., Shanahan, F., Hill, C., Ross, R.P., and O'toole, P.W. (2012). Gut microbiota composition correlates with diet and health in the elderly. *Nature* 488**,** 178-184.
4. Cotsapas, C., and Mitrovic, M. (2018). Genome-wide association studies of multiple sclerosis. *Clin Transl Immunology* 7**,** e1018.
5. Doshi, A., and Chataway, J. (2016). Multiple sclerosis, a treatable disease. *Clin Med (Lond)* 16**,** s53-s59.
6. Dray, S., and Dufour, A.B. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software* 22**,** 1-20.
7. Ezendam, J., De Klerk, A., Gremmer, E.R., and Van Loveren, H. (2008). Effects of Bifidobacterium animalis administered during lactation on allergic and autoimmune responses in rodents. *Clin Exp Immunol* 154**,** 424-431.
8. Filippi, M., Bar-Or, A., Piehl, F., Preziosa, P., Solari, A., Vukusic, S., and Rocca, M.A. (2018). Multiple sclerosis. *Nat Rev Dis Primers* 4**,** 43.
9. Forbes, J.D., Van Domselaar, G., and Bernstein, C.N. (2016). The Gut Microbiota in Immune-Mediated Inflammatory Diseases. *Front Microbiol* 7**,** 1081.
10. Franzosa, E.A., Mciver, L.J., Rahnavard, G., Thompson, L.R., Schirmer, M., Weingart, G., Lipson, K.S., Knight, R., Caporaso, J.G., Segata, N., and Huttenhower, C. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. *Nat Methods* 15**,** 962-968.
11. Ghasemi, N., Razavi, S., and Nikzad, E. (2017). Multiple Sclerosis: Pathogenesis, Symptoms, Diagnoses and Cell-Based Therapy. *Cell J* 19**,** 1-10.
12. Haghikia, A., Jorg, S., Duscha, A., Berg, J., Manzel, A., Waschbisch, A., Hammer, A., Lee, D.H., May, C., Wilck, N., Balogh, A., Ostermann, A.I., Schebb, N.H., Akkad, D.A., Grohme, D.A., Kleinewietfeld, M., Kempa, S., Thone, J., Demir, S., Muller, D.N., Gold, R., and Linker, R.A. (2015). Dietary Fatty Acids Directly Impact Central Nervous System Autoimmunity via the Small Intestine. *Immunity* 43**,** 817-829.
13. Jin, D., Wu, S., Zhang, Y.G., Lu, R., Xia, Y., Dong, H., and Sun, J. (2015). Lack of Vitamin D Receptor Causes Dysbiosis and Changes the Functions of the Murine Intestinal Microbiome. *Clin Ther* 37**,** 996-1009 e1007.
14. Kang, S., Denman, S.E., Morrison, M., Yu, Z., Dore, J., Leclerc, M., and Mcsweeney, C.S. (2010). Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. *Inflamm Bowel Dis* 16**,** 2034-2042.

15. Kaper, J.B., Nataro, J.P., and Mobley, H.L. (2004). Pathogenic Escherichia coli. *Nat Rev Microbiol* 2**,** 123-140.
16. Lee, Y.K., Menezes, J.S., Umesaki, Y., and Mazmanian, S.K. (2011). Proinflammatory T-cell responses to gut microbiota promote experimental autoimmune encephalomyelitis. *Proc Natl Acad Sci U S A* 108 Suppl 1**,** 4615-4622.
17. Lloyd-Price, J., Mahurkar, A., Rahnavard, G., Crabtree, J., Orvis, J., Hall, A.B., Brady, A., Creasy, H.H., Mccracken, C., Giglio, M.G., Mcdonald, D., Franzosa, E.A., Knight, R., White, O., and Huttenhower, C. (2017). Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* 550**,** 61-66.
18. Lopetuso, L.R., Scaldaferri, F., Petito, V., and Gasbarrini, A. (2013). Commensal Clostridia: leading players in the maintenance of gut homeostasis. *Gut Pathog* 5**,** 23.
19. Mangalam, A., Shahi, S.K., Luckey, D., Karau, M., Marietta, E., Luo, N., Choung, R.S., Ju, J., Sompallae, R., Gibson-Corley, K., Patel, R., Rodriguez, M., David, C., Taneja, V., and Murray, J. (2017). Human Gut-Derived Commensal Bacteria Suppress CNS Inflammatory and Demyelinating Disease. *Cell Rep* 20**,** 1269-1277.
20. Mirza, A., and Mao-Draayer, Y. (2017). The gut microbiome and microbial translocation in multiple sclerosis. *Clin Immunol* 183**,** 213-224.
21. Nicaise, A.M., Wagstaff, L.J., Willis, C.M., Paisie, C., Chandok, H., Robson, P., Fossati, V., Williams, A., and Crocker, S.J. (2019). Cellular senescence in progenitor cells contributes to diminished remyelination potential in progressive multiple sclerosis. *Proc Natl Acad Sci U S A* 116**,** 9030-9039.
22. Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., Minchin, P.R., O'hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., and Wagner, H. (2017). vegan: Community Ecology Package. R package version 2.4-3.
23. Pandiyan, P., Bhaskaran, N., Zou, M., Schneider, E., Jayaraman, S., and Huehn, J. (2019). Microbiome Dependent Regulation of Tregs and Th17 Cells in Mucosa. *Front Immunol* 10**,** 426.
24. Rezende, R.M., Oliveira, R.P., Medeiros, S.R., Gomes-Santos, A.C., Alves, A.C., Loli, F.G., Guimaraes, M.A., Amaral, S.S., Da Cunha, A.P., Weiner, H.L., Azevedo, V., Miyoshi, A., and Faria, A.M. (2013). Hsp65-producing Lactococcus lactis prevents experimental autoimmune encephalomyelitis in mice by inducing CD4+LAP+ regulatory T cells. *J Autoimmun* 40**,** 45-57.
25. Rosati, G. (2001). The prevalence of multiple sclerosis in the world: an update. *Neurol Sci* 22**,** 117-139.
26. Saitoh, S., Noda, S., Aiba, Y., Takagi, A., Sakamoto, M., Benno, Y., and Koga, Y. (2002). Bacteroides ovatus as the predominant commensal intestinal microbe causing a systemic antibody response in inflammatory bowel disease. *Clin Diagn Lab Immunol* 9**,** 54-59.
27. Shortt, C., Hasselwander, O., Meynier, A., Nauta, A., Fernandez, E.N., Putz, P., Rowland, I., Swann, J., Turk, J., Vermeiren, J., and Antoine, J.M. (2018). Systematic review of the effects of the intestinal microbiota on selected nutrients and non-nutrients. *Eur J Nutr* 57**,** 25-49.
28. Song, Y., Liu, C., and Finegold, S.M. (2004). Real-time PCR quantitation of clostridia in feces of autistic children. *Appl Environ Microbiol* 70**,** 6459-6465.

29. Team, R.C. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing. *Vienna, Austria*.
30. Truong, D.T., Franzosa, E.A., Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., and Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods* 12**,** 902-903.
31. Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W.H.A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., and Venables, B. (2019). gplots: Various R Programming Tools for Plotting Data. R package version 3.0.1.1.
32. Wexler, H.M. (2007). Bacteroides: the good, the bad, and the nitty-gritty. *Clin Microbiol Rev* 20**,** 593-621.
33. Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York*.
34. Yokote, H., Miyake, S., Croxford, J.L., Oki, S., Mizusawa, H., and Yamamura, T. (2008). NKT cell-dependent amelioration of a mouse model of multiple sclerosis by altering gut flora. *Am J Pathol* 173**,** 1714-1723.
35. Yu, Z., and Morrison, M. (2004). Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* 36**,** 808-812.

Chapter VI

Discussion and future perspectives

Chapter 6 Discussion and future perspectives

The microbiota field has expanded exponentially over the last decade. With reducing cost of sequencing, improvement in sequencing technologies and development of new tools that can not only process large volumes of data but also function computationally faster, various studies have been carried out across the world to understand the relationship of the gut microbiota with the host. Large consortiums such as MetaHIT (Qin et al., 2010) and the Human Microbiome Project (HMP) (Human Microbiome Project, 2012) have obtained valuable information on the microbiota and their functionality. Independent studies and meta-analyses of multiple publicly available datasets have shown how the microbiota is shaped across ethnicity, genetics, age-demographics, diseases, clinical factors, lifestyle, diet and other social and economic factors (Duvallet et al., 2017;Jackson et al., 2018;Almeida et al., 2019;Pasolli et al., 2019). These large studies shed light into the malleable nature of the gut microbiota which is important for not only uncovering confounders but also ensuring proper microbiota study design and discovering biologically relevant signatures. This lays the foundation for future clinical research that may involve studies of probiotics, faecal microbiota transplant, or investigation of the mechanistic role of disease associated marker taxa. Understanding these links along with the effect of various covariates will allow researchers to address the issues with reproducibility of results also. The reproducibility concerns have led to the establishment of Microbiome Quality Control project (Sinha et al., 2015) which provides guidelines for microbiota data analysis and researches effectors of microbiota alterations.

A key aspect of microbiota studies is to identify taxa and microbial products which may contribute as risk factors for immune-associated diseases. Such characteristic features would have the potential as markers for predicting onset or

diagnosis of immune-associated conditions. The changes in the microbiota profiles may also be associated with treatment and disease activity which would serve as indicators for remission and efficacy of treatment. Insights from various microbiota research has established their prominent role in immune-homeostasis, production of nutrients and metabolites crucial for the health of the host (Cianci et al., 2018;Valdes et al., 2018). Numerous studies have identified various gut microbes with pro- and anti-inflammatory activities (Cianci et al., 2018). Presence or absence of commensal microbiota has been linked with metabolic abnormality in the host, increased abundance of pathobionts or development of immune-associated disorders (Stecher, 2015). However, microbiota-based human studies with conditions like osteoporosis and different types of arthritis that include the investigation of the effect of immune-modulatory and disease modifying non-antibiotic medications are limited. This is discussed thoroughly in the thesis where the respective chapters explore the relationship between the gut microbes and various immune-associated diseases. It generates a knowledge base of the microbiota as a risk factor and acknowledges the therapeutic potential of the microbiota in modulating the immune diseases.

Chapter 2 of the thesis is an observational study that explores the gut-bone axis. This study is the first and one of the largest human study that identifies the changes in the gut microbial community in medically classified osteopenia and osteoporosis subjects in comparison to age- and gender-match healthy subjects. The study also includes extensive analysis of dietary, clinical, and anthropometric measures and accounts for the association of various covariates with microbiota. These various covariates showed similar trends and effect size as reported in literature. After identifying and adjusting for various confounders, the study identified a specific set of genera that were associated with bone mineral density (BMD) measure and remained

significantly associated with BMD even after removing the effect of various medications and BMI. However, it was noted the *Clostridium XIVa* and *Veillonella* remained associated with both BMI and BMD. The identified taxa have been reported individually to be associated in various bone and immune-associated studies, thus supporting the results of the study. This study benefits from large comparable sample size per group, use of robust and rigorous statistical approaches and extensive meta-data investigation. The study has been published in Rheumatology in which an editorial has also been written on the manuscript which is a recognition of the study (Aurora, 2019).

Our results are divergent from a recent study by Li *et al*. on a Chinese cohort of 102 samples that reported various taxonomic and functional alterations to be associated with BMD measures. However, that study segregated their data based on median BMD value and not clinically defined criteria. It also lacked adjustment for multiple testing except for the functional analysis (Li et al., 2019). Wang *et al.* published a preprint recently that investigated the relationship between BMD and gut microbiota using shotgun whole genome metagenomics (mWGS) sequence data on a large cohort of 361 elderly Chinese women. The associated species identified in this study belonged to the significant genera identified in our study validating the results of our research (Wang et al., 2019b). Overall, the current study from chapter 2 improves the understanding of gut-bone axis which could lead to development of microbiota-based diagnostic/therapeutic approach. This would have tremendous applications in early detection and/or in providing intervention to the at-risk elderly population through microbiota modulation, thereby improving the quality of life and reducing economic burden.

Chapter 3 of the thesis investigates the changes in the gut microbiota dynamics in three different arthritic conditions: ankylosing spondylitis (AS), rheumatoid arthritis (RA) and psoriatic arthritis (PSA). This is studied by sampling the subjects across three time-points during the course of biologics treatment. As these conditions are extraintestinal, to evaluate the degree of the gut microbiota association with inflammation, IBD samples were also included as an outgroup. The study also benefits with the inclusion of established long-term biologics treated groups comprising of different samples with same arthritic conditions. This study is unique as it is the first explore the link between biologics treatment and microbiota from arthritis and Inflammatory Bowel disease (IBD) along with extensive clinical, dietary, serum and anthropometric information. Unlike the previous chapter, this study was carried out with mWGS sequence data allowing investigation of species level taxonomic and the functional profiles. The time-point sampling and the inclusion of established treated samples allowed us to track the changes in the gut microbiota during the course of treatment and control for various covariates associated with inter-individual differences like dietary habits, lifestyle etc. The study observed that the long-term biologics treated arthritic groups did not show significant difference in diversity from healthy controls which could be an indicator of similar gut microbiota composition. Interestingly, it was observed that samples from RA and PSA individuals at baseline (T0) were not significantly different from controls which might be partly due to methotrexate treatment that was absent is AS group. This suggest that the microbiota alterations are associated with immuno-suppressive treatments. IBD, in agreement with previous literature, showed significantly reduced α-diversity and altered β-diversity (Halfvarson et al., 2017). Various healthy controls associated taxa were also depleted in the IBD samples. The presence of both control and disease associated taxa

in different established arthritic groups and their absence in T0 samples may indicate partial restoration. The validation with independent RA dataset where similar signals were observed further confirms our results (Zhang et al., 2015). The correlation between different taxonomic signatures and clinical data were also consistent with literature (Wexler, 2007;Lopetuso et al., 2013). The lack of distinct functional signature may be indicator of redundancy in functional potential. Despite the sample size limitation in the study, the identification of established signatures and the presence of healthy associated taxa in the long-term biologics treated samples justifies the hypothesis that the altered gut microbiota composition is partially restored during remission upon treatment with immunosuppressants.

Understanding the dynamics of gut microbes in immune-mediated disorders involves uncovering taxonomic markers and bacterial metabolites that can trigger or suppress immune responses in the intestine. Various immune-associated inflammatory conditions such as different types of arthritis, IBD and multiple sclerosis (MS) have identified gut microbiota as an external risk factor (Clemente et al., 2018). The relationship between the gut microbiota and immune cells can be bi-directional where the immune cells modulate the bacterial population in the gut and the microbiota influences immune cell activities. This also raises the question of how non-antibiotic medications such as immunosuppressants affects this relationship between gut microbiota and inflammation (Busquets et al., 2015;Zhang et al., 2015;Bazin et al., 2018). This is important as it may shed light into the dynamics of gut microbiota in different immune disorders as studies have shown a partial restoration of the gut microbiota towards healthy controls during remission stage.

Similarly, Chapter V of the thesis studied the gut microbiota signature in different multiple sclerosis (MS) phenotypes. The findings in this chapter also reflect

the depletion of healthy associated bacterial species and presence of potential pro-inflammatory associated and pathogenic taxa. It also indicates that different subtypes of MS may share similar microbiota signature as evident from primary-progressive (PPMS) and secondary progressive (SPMS) samples. While the immune-associated characteristic taxonomic signatures were observed, the low number of MS samples in the study is a limitation to the robustness of the signature. The lack of meta-data also limits the understanding of the microbiota-host interactions in this case as microbiota is affected by various covariates. Despite the limitation, the microbiota profiles show an altered composition, which is distinct from both young and elderly healthy population. Reproduction of taxonomic alterations as published in various independent studies reaffirms our study validating the biological association of the gut microbiota in MS which is distinct from healthy ageing. These altered microbiota profiles can be potentially exploited for development of early detection and disease activity monitoring system. Furthermore, there is potential for development of microbiota based live biotherapeutics from key taxonomic markers with disease modulatory properties.

Chapter 4 of the thesis was focused on development of a novel methodology for functional inference from amplicon dataset. The functional potential of the gut microbiota was necessary for understanding the microbiota interaction with the immune system. Data generated from bacterial transcripts, proteins and metabolites can allow assessment of the activity of the microbes in the gut. However, the generation of such data can be limited by various technical challenges. Alternatively, the identification of the microbial genes and estimation of the functional potential is achieved with mWGS sequencing. The use of amplicon sequencing remains popular due to its low cost and computational necessity but lacks functional information. To

extrapolate the functional capabilities from amplicon data, inference tools have been developed (Langille et al., 2013;Asshauer et al., 2015;Iwai et al., 2016). However, these tools are limited due to their reliance on functionally annotated taxa.

Our methodology: **I**nference of **P**athways from **C**o-**V**ariance analysis (IPCO), implements a double co-inertia approach for inferring functionality of an amplicon dataset. The advantage of IPCO is that it is not limited to the use of one particular type of annotation or reliance on specific databases. We have shown that IPCO works well with both KEGG and MetaCyc annotation. IPCO can also be easily implemented using an in-house reference dataset instead of the provided references. This translates to the flexibility of IPCO, which allows better comparison with other datasets avoiding the requirement of re-processing the data towards a specific type of annotation. Another highlight of IPCO is that the predicted values of IPCO shows the same directionality for various features (pathways) as observed in mWGS dataset which is assumed to reflect of the actual biological signals. This signal is reproduced to the greatest extent using IPCO. This signature is not observed using any of the other publicly available methods. This highlights the lack of predicted pathways to mWGS pathways abundance correlation across various samples in the available methods. This would result is spurious identification of significantly differential functionality, incorrect assumption of the directionality of abundance, thus leading to further inconclusive observations. The study also highlighted that concordance between functional and taxonomic profiles obtained from different sites is different. This is partially reflective of the level of unannotated information which requires further investigation. IPCO is implemented in R, which is widely used for statistical analyses. Various other tools such as DADA2 (Callahan et al., 2016), phyloseq (McMurdie and Holmes, 2013) that allows processing 16S data are completely implemented in R. IPCO is made available

as a preprint, which has been viewed nearly 300 times as full text as of November 2019.

Future improvements of IPCO would include better inference from other environments and body sites besides stool samples, assigning a confidence value to the prediction, and identification of the taxonomic contribution similar to MelonnPan (Mallick et al., 2019). The current version of IPCO is also limited to the KEGG version of HUMAnN1 (Abubucker et al., 2012) and could be updated in future. Superclass information for the MetaCyc pathways would also be provided in the future. To improve the usability of IPCO, scripts/tutorials for processing the output of IPCO and downstream analysis can be added to help naive users. PICRUSt2 (Douglas et al., 2019) and Tax4Fun2 (Wemheuer et al., 2018) have also been made available as preprint version. IPCO would also benefit from benchmarking against these tools once published.

Overall, research carried out as part of the thesis improves the understanding of the gut microbiota in immune diseases. Amplicon and mWGS sequencing provided the necessary information for the current research. Despite this, there is a need for different multi-omics profiles to further elucidate the microbiota-host interactions (Wang et al., 2019a). Schirmer *et al.* investigated a IBD cohort with metatranscriptomics data and have shown the strengths of analysing the functional activity (Schirmer et al., 2018). The group identified different taxa which showed differential abundance in IBD samples but lacked any functional activity. It also highlighted that despite the similar abundances of certain taxa their functional activities were different in the patients through the duration of the disease. Similarly shifts in biochemical and functional properties was also noted by Lloyd-Price *et al.* (Lloyd-Price et al., 2019). Bacterial transcripts can have a short life span which could

be challenge if there is a long time gap due to transport, storage, extraction and sequencing (Allaband et al., 2019). Bacterial metabolites profiles are also very informative as metabolites such as short chain fatty acids (SCFAs) are associated with healthy fibre associated microbiota and gut-immune homeostasis (Makki et al., 2018). Secondary bile acids have also been linked with colon cancer which further highlights the necessary for evaluating the metabolite profiles (Ocvirk and O'Keefe, 2017). Recently AGORA has been published which provides metabolite capacity for 773 gut bacteria (Magnusdottir et al., 2017). Although, such resources are helpful in inferring metabolite profiles, these may not be reflective of the actual profiles as many unknown species are also present in the gut.

The importance of various clinical, biochemical, dietary, and other meta-data cannot be overstated. This is observed in Chapter 5 of the thesis where lack of clinical information is a limitation in the study. Chapter 2 and 3 in the thesis reflects a better comprehensive study due to the availability of various meta-data. The relationship between microbiota and various covariates is stressed in the publications by Zhernakova *et al.* and Falony *et al.* (Falony et al., 2016;Zhernakova et al., 2016). The use of amplicon sequencing to generate 16S datasets are mostly reported down to genus level. With newer tools such as DADA2, Deblur (Amir et al., 2017) and species level classifiers like SPINGO (Allard et al., 2015), working with sub-genus level datasets would allow better understanding of the gut microbiota as different species from the same genus can harbour different genes or may have pro- or anti-inflammatory functions. E.g. *Clostridium* and *Bacteroides* (Wexler, 2007;Lopetuso et al., 2013). Increased taxonomic resolution down to strain levels is also necessary as different strains of the same species can harbour different functional capacity like pathogenic capability e.g. *Escherichia coli* (Tenaillon et al., 2010).

The work carried out as part of thesis provides evidences of microbiota-host association in relation to immune-associated conditions. Nevertheless, the studies are not without limitations. In chapter 2, the study identifies confounding effect of obesity measures that are linked with BMD. This makes it difficult to isolate the signature between microbiota and BMD due to the co-variance between BMD and BMI. The large differences in microbiota profiles seen when comparing below weight, normal BMI and obese individuals makes addressing this confounder a particularly important concern. Although statistical modelling was utilised to remove this effect, the possibility remains that these differences may be due in part to non-BMD differences between groups. To fully account for BMI, it would be necessary to confront this confounder at the recruitment stage of the study and match the groups for BMI as well as the other confounders such as gender, age and frailty. This should be the goal of any future observational study looking at the microbiota and BMD.

Also, from chapter 2, 3, and 5, which investigates the association between the microbiota and various diseases, it is evident that an experimental design that follows the subjects over the time would be beneficial. This would allow a better understanding of how habitual, diet, clinical and medication factors and other changes in the subjects will influence and be influenced by the microbiota. A longitudinal study would allow better accounting for the stability and variations in the microbiota population structure during disease pathogenesis. In a clinical setting, accounting for medications is necessary when investigating the microbiota signature across and during disease activity and treatment duration. When testing for the effect of disease modifying medications with respect to microbiota, it would be statistically beneficial to include a placebo group. The placebo group would allow us to investigate the effect disease pathogenesis on the microbiota in absence of medications. This would allow

better understanding of specific microbiota signatures associated with disease pathogenesis and the effect of medications on the microbiota. However, a number of ethical considerations would need to be explored before such a study could proceed.

While investigating the microbiota population from the human gut, most studies focus on the bacterial component as archaea species are sparsely populated (usually below detection threshold). Similarly, due to lack of universally common region in viruses; bacteriophages, which dominates the viral component in the human gut (Sutton and Hill, 2019), cannot be captured using amplicon sequencing. Although mWGS approaches can capture all the genomic content, lysogenic stage (viral genome integrated into host genome without killing the host) of the bacteriophages makes it complicated to separate the viral and bacterial genome. Hence, a separate phage isolation and sequencing protocol is followed for studying bacteriophages.

The focus on the bacterial content of the microbiota overlook the presence of an organism(s) that perform key functions in health and disease. For instance, species of archaea are able of metabolise TMA preventing its absorption in the intestine (Nkamga et al., 2017). Archaea utilises hydrogen from the gut and converts it into methane. Removal of hydrogen indirectly allows more ATP production by the gut bacteria promoting their growth (Nkamga et al., 2017). Certain archaeal species can convert heavy metal into methylated products which exhibits toxic effect towards gut bacteria and host (Nkamga et al., 2017). These capabilities of archaea show that they are critical members of the gut microflora and their metabolic activities have a direct and/or indirect impact on bacterial population and the host.

Similarly, bacteriophages also play an important role in regulating bacterial population. Without accounting for bacteriophages, the understanding of microbiota population dynamics would remain incomplete. The bacteriophages are responsible

for horizontal gene transfer, infection and predation of bacterial host, which also results in evolutionary pressure on the bacterial host (Sutton and Hill, 2019). Bacteriophages are highly specific and predation of susceptible bacterial strains allows selection of phage resistant strains. Such changes result in mutation of viral genomes and selection of bacteriophages with infectivity leading to a continuous cycle of changes in population dynamics between bacteriophages and its host (Weinbauer, 2004; Scanlan, 2017). Recently Clooney *et al.* reported that the bacteriophage population showed concordance with the bacterial community and accounting for both bacteriophage and bacterial composition allowed better predictive power between IBD and control samples thereby highlighting the importance and necessity of accounting for phage population while studying the microbiota (Clooney et al., 2019).

In conclusion, the research on understanding the relationship of gut microbiota in various immune disorders carried out as part of this thesis contributes significantly towards the microbiota knowledge and the development of necessary methodology for microbiota studies. From chapters 2, 3 and 5, the novel findings add significantly to the biological understanding to the role of microbiota in inflammation and vice-versa. The research further reaffirms the relationship with various meta-data such as medication, biochemical properties, and dietary factors. The replication of previously reported signature not only validates the biological signal reported in this thesis but also stresses the importance of gut microbiota as a risk factor in different disorders. The reduced abundance of the health associated taxa such as those involved in SCFA production, modulation of T-regulatory cells is an indicator that these taxa are necessary for gut homeostasis and their depletion can trigger altered immune responses. The presence of commensal taxa is also necessary for controlling the abundances of pathobionts, which may further aggravate inflammation if allowed to

proliferate. The observations from the various studies carried out as part of the thesis provides evidence for various taxonomic markers that may have the potential for diagnostic tool and/or predicting remission in immune conditions. The research carried out in this thesis provides many conclusive answers but also raises new challenges for future studies. Such studies could involve investigation of the identified microbiota alterations for their mechanistic role in immune-disorder and the effect of immunosuppressive medications. The study can be further expanded to investigate the potential health-promoting capabilities of the taxa that are associated with beneficial properties. These would hopefully guide the field towards a better understanding of the mechanistic link between gut microbiota and the host and may lead to development of novel therapeutics.

References

1. Abubucker, S., Segata, N., Goll, J., Schubert, A.M., Izard, J., Cantarel, B.L., Rodriguez-Mueller, B., Zucker, J., Thiagarajan, M., Henrissat, B., White, O., Kelley, S.T., Methe, B., Schloss, P.D., Gevers, D., Mitreva, M., and Huttenhower, C. (2012). Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput Biol* 8**,** e1002358.
2. Allaband, C., Mcdonald, D., Vazquez-Baeza, Y., Minich, J.J., Tripathi, A., Brenner, D.A., Loomba, R., Smarr, L., Sandborn, W.J., Schnabl, B., Dorrestein, P., Zarrinpar, A., and Knight, R. (2019). Microbiome 101: Studying, Analyzing, and Interpreting Gut Microbiome Data for Clinicians. *Clin Gastroenterol Hepatol* 17**,** 218-230.
3. Allard, G., Ryan, F.J., Jeffery, I.B., and Claesson, M.J. (2015). SPINGO: a rapid species-classifier for microbial amplicon sequences. *BMC Bioinformatics* 16**,** 324.
4. Almeida, A., Mitchell, A.L., Boland, M., Forster, S.C., Gloor, G.B., Tarkowska, A., Lawley, T.D., and Finn, R.D. (2019). A new genomic blueprint of the human gut microbiota. *Nature* 568**,** 499-504.
5. Amir, A., Mcdonald, D., Navas-Molina, J.A., Kopylova, E., Morton, J.T., Zech Xu, Z., Kightley, E.P., Thompson, L.R., Hyde, E.R., Gonzalez, A., and Knight, R. (2017). Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *mSystems* 2.
6. Asshauer, K.P., Wemheuer, B., Daniel, R., and Meinicke, P. (2015). Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics* 31**,** 2882-2884.
7. Aurora, R. (2019). Confounding factors in the effect of gut microbiota on bone density. *Rheumatology (Oxford)*.
8. Bazin, T., Hooks, K.B., Barnetche, T., Truchetet, M.E., Enaud, R., Richez, C., Dougados, M., Hubert, C., Barre, A., Nikolski, M., and Schaeverbeke, T. (2018). Microbiota Composition May Predict Anti-Tnf Alpha Response in Spondyloarthritis Patients: an Exploratory Study. *Sci Rep* 8**,** 5446.
9. Busquets, D., Mas-De-Xaxars, T., Lopez-Siles, M., Martinez-Medina, M., Bahi, A., Sabat, M., Louvriex, R., Miquel-Cusachs, J.O., Garcia-Gil, J.L., and Aldeguer, X. (2015). Anti-tumour Necrosis Factor Treatment with Adalimumab Induces Changes in the Microbiota of Crohn's Disease. *J Crohns Colitis* 9**,** 899-906.
10. Callahan, B.J., Mcmurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 13**,** 581-583.
11. Cianci, R., Pagliari, D., Piccirillo, C.A., Fritz, J.H., and Gambassi, G. (2018). The Microbiota and Immune System Crosstalk in Health and Disease. *Mediators Inflamm* 2018**,** 2912539.
12. Clemente, J.C., Manasson, J., and Scher, J.U. (2018). The role of the gut microbiome in systemic inflammatory disease. *BMJ* 360**,** j5145.
13. Clooney, A.G., Sutton, T.D.S., Shkoporov, A.N., Holohan, R.K., Daly, K.M., O'regan, O., Ryan, F.J., Draper, L.A., Plevy, S.E., Ross, R.P., and Hill, C. (2019). Whole-Virome Analysis Sheds Light on Viral Dark Matter in Inflammatory Bowel Disease. *Cell Host Microbe* 26**,** 764-778 e765.

14. Douglas, G.M., Maffei, V.J., Zaneveld, J., Yurgel, S.N., Brown, J.R., Taylor, C.M., Huttenhower, C., and Langille, M.G.I. (2019). PICRUSt2: An improved and extensible approach for metagenome inference. *bioRxiv*, 672295.
15. Duvallet, C., Gibbons, S.M., Gurry, T., Irizarry, R.A., and Alm, E.J. (2017). Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat Commun* 8, 1784.
16. Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K., Kurilshikov, A., Bonder, M.J., Valles-Colomer, M., Vandeputte, D., Tito, R.Y., Chaffron, S., Rymenans, L., Verspecht, C., De Sutter, L., Lima-Mendez, G., D'hoe, K., Jonckheere, K., Homola, D., Garcia, R., Tigchelaar, E.F., Eeckhaudt, L., Fu, J., Henckaerts, L., Zhernakova, A., Wijmenga, C., and Raes, J. (2016). Population-level analysis of gut microbiome variation. *Science* 352, 560-564.
17. Halfvarson, J., Brislawn, C.J., Lamendella, R., Vazquez-Baeza, Y., Walters, W.A., Bramer, L.M., D'amato, M., Bonfiglio, F., Mcdonald, D., Gonzalez, A., Mcclure, E.E., Dunklebarger, M.F., Knight, R., and Jansson, J.K. (2017). Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat Microbiol* 2, 17004.
18. Human Microbiome Project, C. (2012). Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207-214.
19. Iwai, S., Weinmaier, T., Schmidt, B.L., Albertson, D.G., Poloso, N.J., Dabbagh, K., and Desantis, T.Z. (2016). Piphillin: Improved Prediction of Metagenomic Content by Direct Inference from Human Microbiomes. *PLoS One* 11, e0166104.
20. Jackson, M.A., Verdi, S., Maxan, M.E., Shin, C.M., Zierer, J., Bowyer, R.C.E., Martin, T., Williams, F.M.K., Menni, C., Bell, J.T., Spector, T.D., and Steves, C.J. (2018). Gut microbiota associations with common diseases and prescription medications in a population-based cohort. *Nat Commun* 9, 2655.
21. Langille, M.G., Zaneveld, J., Caporaso, J.G., Mcdonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Vega Thurber, R.L., Knight, R., Beiko, R.G., and Huttenhower, C. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* 31, 814-821.
22. Li, C., Huang, Q., Yang, R., Dai, Y., Zeng, Y., Tao, L., Li, X., Zeng, J., and Wang, Q. (2019). Gut microbiota composition and bone mineral loss-epidemiologic evidence from individuals in Wuhan, China. *Osteoporos Int* 30, 1003-1013.
23. Lloyd-Price, J., Arze, C., Ananthakrishnan, A.N., Schirmer, M., Avila-Pacheco, J., Poon, T.W., Andrews, E., Ajami, N.J., Bonham, K.S., Brislawn, C.J., Casero, D., Courtney, H., Gonzalez, A., Graeber, T.G., Hall, A.B., Lake, K., Landers, C.J., Mallick, H., Plichta, D.R., Prasad, M., Rahnavard, G., Sauk, J., Shungin, D., Vazquez-Baeza, Y., White, R.A., 3rd, Investigators, I., Braun, J., Denson, L.A., Jansson, J.K., Knight, R., Kugathasan, S., Mcgovern, D.P.B., Petrosino, J.F., Stappenbeck, T.S., Winter, H.S., Clish, C.B., Franzosa, E.A., Vlamakis, H., Xavier, R.J., and Huttenhower, C. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* 569, 655-662.
24. Lopetuso, L.R., Scaldaferri, F., Petito, V., and Gasbarrini, A. (2013). Commensal Clostridia: leading players in the maintenance of gut homeostasis. *Gut Pathog* 5, 23.

25. Magnusdottir, S., Heinken, A., Kutt, L., Ravcheev, D.A., Bauer, E., Noronha, A., Greenhalgh, K., Jager, C., Baginska, J., Wilmes, P., Fleming, R.M., and Thiele, I. (2017). Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat Biotechnol* 35**,** 81-89.
26. Makki, K., Deehan, E.C., Walter, J., and Backhed, F. (2018). The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease. *Cell Host Microbe* 23**,** 705-715.
27. Mallick, H., Franzosa, E.A., Mclver, L.J., Banerjee, S., Sirota-Madi, A., Kostic, A.D., Clish, C.B., Vlamakis, H., Xavier, R.J., and Huttenhower, C. (2019). Predictive metabolomic profiling of microbial communities using amplicon or metagenomic sequences. *Nat Commun* 10**,** 3136.
28. Mcmurdie, P.J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8**,** e61217.
29. Nkamga, V.D., Henrissat, B., and Drancourt, M. (2017). Archaea: Essential inhabitants of the human digestive microbiota. *Human Microbiome Journal* 3**,** 1-8.
30. Ocvirk, S., and O'keefe, S.J. (2017). Influence of Bile Acids on Colorectal Cancer Risk: Potential Mechanisms Mediated by Diet - Gut Microbiota Interactions. *Curr Nutr Rep* 6**,** 315-322.
31. Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P., Collado, M.C., Rice, B.L., Dulong, C., Morgan, X.C., Golden, C.D., Quince, C., Huttenhower, C., and Segata, N. (2019). Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* 176**,** 649-662 e620.
32. Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D.R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H.B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Dore, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Meta, H.I.T.C., Bork, P., Ehrlich, S.D., and Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464**,** 59-65.
33. Scanlan, P.D. (2017). Bacteria-Bacteriophage Coevolution in the Human Gut: Implications for Microbial Diversity and Functionality. *Trends Microbiol* 25**,** 614-623.
34. Schirmer, M., Franzosa, E.A., Lloyd-Price, J., Mciver, L.J., Schwager, R., Poon, T.W., Ananthakrishnan, A.N., Andrews, E., Barron, G., Lake, K., Prasad, M., Sauk, J., Stevens, B., Wilson, R.G., Braun, J., Denson, L.A., Kugathasan, S., Mcgovern, D.P.B., Vlamakis, H., Xavier, R.J., and Huttenhower, C. (2018). Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat Microbiol* 3**,** 337-346.
35. Sinha, R., Abnet, C.C., White, O., Knight, R., and Huttenhower, C. (2015). The microbiome quality control project: baseline study design and future directions. *Genome Biol* 16**,** 276.

36. Stecher, B. (2015). The Roles of Inflammation, Nutrient Availability and the Commensal Microbiota in Enteric Pathogen Infection. *Microbiol Spectr* 3.
37. Sutton, T.D.S., and Hill, C. (2019). Gut Bacteriophage: Current Understanding and Challenges. *Front Endocrinol (Lausanne)* 10**,** 784.
38. Tenaillon, O., Skurnik, D., Picard, B., and Denamur, E. (2010). The population genetics of commensal Escherichia coli. *Nat Rev Microbiol* 8**,** 207-217.
39. Valdes, A.M., Walter, J., Segal, E., and Spector, T.D. (2018). Role of the gut microbiota in nutrition and health. *BMJ* 361**,** k2179.
40. Wang, Q., Wang, K., Wu, W., Giannoulatou, E., Ho, J.W.K., and Li, L. (2019a). Host and microbiome multi-omics integration: applications and methodologies. *Biophys Rev* 11**,** 55-65.
41. Wang, Q., Zhao, H., Sun, Q., Li, X., Chen, J., Wang, Z., Ju, Y., Jie, Z., Guo, R., Liang, Y., Sun, X., Zheng, H., Zhang, T., Xiao, L., Xu, X., Yang, H., Peng, S., and Jia, H. (2019b). Shotgun Metagenomics of 361 elderly women reveals gut microbiome change in bone mass loss. *bioRxiv***,** 679985.
42. Weinbauer, M.G. (2004). Ecology of prokaryotic viruses. *FEMS Microbiol Rev* 28**,** 127-181.
43. Wemheuer, F., Taylor, J.A., Daniel, R., Johnston, E., Meinicke, P., Thomas, T., and Wemheuer, B. (2018). Tax4Fun2: a R-based tool for the rapid prediction of habitat-specific functional profiles and functional redundancy based on 16S rRNA gene marker gene sequences. *bioRxiv***,** 490037.
44. Wexler, H.M. (2007). Bacteroides: the good, the bad, and the nitty-gritty. *Clin Microbiol Rev* 20**,** 593-621.
45. Zhang, X., Zhang, D., Jia, H., Feng, Q., Wang, D., Liang, D., Wu, X., Li, J., Tang, L., Li, Y., Lan, Z., Chen, B., Li, Y., Zhong, H., Xie, H., Jie, Z., Chen, W., Tang, S., Xu, X., Wang, X., Cai, X., Liu, S., Xia, Y., Li, J., Qiao, X., Al-Aama, J.Y., Chen, H., Wang, L., Wu, Q.J., Zhang, F., Zheng, W., Li, Y., Zhang, M., Luo, G., Xue, W., Xiao, L., Li, J., Chen, W., Xu, X., Yin, Y., Yang, H., Wang, J., Kristiansen, K., Liu, L., Li, T., Huang, Q., Li, Y., and Wang, J. (2015). The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nat Med* 21**,** 895-905.
46. Zhernakova, A., Kurilshikov, A., Bonder, M.J., Tigchelaar, E.F., Schirmer, M., Vatanen, T., Mujagic, Z., Vila, A.V., Falony, G., Vieira-Silva, S., Wang, J., Imhann, F., Brandsma, E., Jankipersadsing, S.A., Joossens, M., Cenit, M.C., Deelen, P., Swertz, M.A., Lifelines Cohort, S., Weersma, R.K., Feskens, E.J., Netea, M.G., Gevers, D., Jonkers, D., Franke, L., Aulchenko, Y.S., Huttenhower, C., Raes, J., Hofker, M.H., Xavier, R.J., Wijmenga, C., and Fu, J. (2016). Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* 352**,** 565-569.

Acknowledgements

My PhD journey is at its end and I would like to take this opportunity to thank everyone without whom I could not have completed this journey. To start at the beginning, I would like to thank Dr. Ian Jeffery and Prof. Paul O'Toole who gave me this exciting PhD opportunity. I am eternally grateful and would like to sincerely thank Ian for considering me as his first PhD student. I learnt a lot from him, and I appreciate all the support that I got throughout my PhD. I would also like to extend my sincere gratitude towards my co-supervisor Paul for his expert opinion and straightforward critical feedbacks for which I am deeply thankful. Both of my supervisors have helped me greatly to develop a scientific temper which has help me grow as a researcher. By working with Paul's lab group, I also got to interact with a highly diverse group of scientists and experience a world class collaborative atmosphere. I would also like to thank all the co-authors and the staffs who worked on the various projects that I have presented in this thesis. Without their help, my work would have remained incomplete. I am grateful to Science Foundation Ireland for supporting my PhD and APC Microbiome Ireland for the infrastructure, administrative and technical support. Working in APC Microbiome Ireland has not only exposed me with world-class research but have also provided opportunities to develop skills beyond research, encouraging holistic development.

Moving on to my colleagues and friends, A special thanks to my close friends in Ireland Adam, David, Feargal, Hugh, Maurice, Ross, Sidney, Tom and by extension their friends and girlfriends. They made sure that I did not feel like an outsider and helped me a lot through all ups and down during my PhD and made my life in Ireland very pleasant with lots of fond memories. I would also like to thank all the current and previous members of office 4.11 and 5.07 with whom I shared an office space. I am thankful to all the lab members from O'Toole's group Alexendra Ntemiri, Ana

Almeida, Burkhardt Flemer, Celine Ribiere, Claire Hegarty, Edel Cormac, Elisa Salvetti, Klára Vlčková, Marta Neto, Marta Parez, Paola Pellanda, Tam Tran, Tarini Ghosh, Werner Frei, and Emily Power from Claesson's group and Janaki Velmurugan from Nally's group who are great people to talk and work with in APC Microbiome Ireland. Speaking of my experiences in Ireland, I would also like to thank my MSc friends for making my first year in Ireland really enjoyable. Aoife, Christy, David, Eibhlin, Kevin, Laura, Michelle, Nicola, Rachel, Stephan Kiniry and Stephan Murphy who were the first friends that I made coming to Ireland.

Looking back there are a lot of people in India who always believed in me, guided, and supported me. A big thanks to the fantastic bunch of friends Anuj, Ashok, Kamlesh, Ravinder, Richa, Sarvesh, Sonakshi, Suparna, Vinay, and Zaid for those delightful years in Tamil Nadu and for sharing the same enthusiasm for research as me and for motivating me. A special thanks to Neelam for being a close confidant and keeping me on the right track. I would like to express my deepest gratitude to Archana, Bipin, Kushal, Prashant, Radheshyam, Sandhya, Shymanta, Sumitra, and Sushana from Bangalore whom I consider as my "second family". I started my college life in a foreign city with some of you and met the rest of you along the way and those days were the best. I learnt to live life to the fullest from you all. I would like to mention my mentors Dr. Antoney P.U and Dr. Suma S. from my undergraduate years and thank them for all the help and advices they gave which have helped me to be where I am today. To my few school friends Karuna, Manas, Reeha, Ruhismita and Samiron, I thank you all for always being there for me back home.

I would like to specially thank Bonani and Ipra who always encouraged, supported, and sometimes "bullied" me. Their love and silly antics made me more confident in life. To two of my best friends Arunabh and Pankaj, who are always there

with me through thick and thin and helping me whenever I needed help even from the good old school days. I have definitely missed a few people, but I am grateful to them also. I am thankful to my relatives and family members for all the encouragement, support, and love throughout my life. At the end, I would like to dedicate this thesis to my parents and thank them for taking care of me, guiding, encouraging, and supporting me in my life decisions.